STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 114217

To: Sean McGarry Location: REM-2C18 Art Unit: 1635 Thursday, February 19, 2004 Case Serial Number: 10/026341	From: Beverly Shears Location: Remsen Bldg. RM 1A54 Phone: 571-272-2528 beverly.shears@uspto.gov

Search Notes

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____

Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevent citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevent claim(s).

STAFF USE ONLY

Date completed: 02-19-04

Searcher: Bessy C 2528

Terminal time: 20

Elapsed time: _____

CPU time: _____

Total time: 25

Number of Searches: _____

Number of Databases: 1

Search Site

_____ STIC
_____ CM-1
_____ Pre-S

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors

_____ IG
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
✓ Other CGN

PTO-1590 (9-90)

McGarry
10/026341
Seq. IDs 1 & 2

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 11:13:25 ; Search time 1719.07 Seconds
(without alignments)
5592.433 Million cell updates/sec

Title: US-10-026-341A-1
Perfect score: 235
Sequence: 1 atggctggcagatcatctct.....ccacagttccagaccgttga 235

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		235	100.0	235	6	AX526280	AX526280 Sequence
2		235	100.0	2655	9	HUMTFSP1	J03133 Human trans
3		235	100.0	2806	9	BC043224	BC043224 Homo sapi
4		235	100.0	2913	9	AF252284	AF252284 Homo sapi
5		235	100.0	4300	9	AB039286	AB039286 Homo sapi
6	c	235	100.0	73433	9	AC068889	AC068889 Homo sapi
7		222.2	94.6	2475	10	RATSP1	D12768 Rattus norv
8		222.2	94.6	4092	10	AB077988	AB077988 Rattus no
9		222.2	94.6	206520	2	AC109743	AC109743 Rattus no
10		222.2	94.6	254686	2	AC097309	AC097309 Rattus no
11		221	94.0	166697	2	AC021103	AC021103 Homo sapi
12		220.6	93.9	4128	10	AF022363	AF022363 Mus muscu
13		192.6	82.0	586	11	G92513	G92513 S208P6120RE
14		192.6	82.0	3741	10	AF062566	AF062566 Mus muscu
15	c	192.6	82.0	138860	2	AC055703	AC055703 Mus muscu
16		192.6	82.0	205187	2	AC137156	AC137156 Mus muscu
17		84.2	35.8	2662	5	GGA317960	AJ317960 Gallus ga
18		72.4	30.8	210390	2	AC110247	AC110247 Mus muscu
19		72.4	30.8	243137	2	AC119846	AC119846 Mus muscu
20		36.4	15.5	4954	3	AF213376	AF213376 Drosophil
21	c	36.4	15.5	13505	2	AC015161	AC015161 Drosophil
22	c	36.4	15.5	54695	2	AC010004	AC010004 Drosophil
23	c	36.4	15.5	185497	3	AC010563	AC010563 Drosophil
24	c	36.4	15.5	279088	3	AE003518	AE003518 Drosophil
25		36.2	15.4	225727	2	AC117607	AC117607 Mus muscu
26		35	14.9	184735	9	AC016180	AC016180 Homo sapi
27		34.6	14.7	930	9	AY128662	AY128662 Homo sapi
28		34.6	14.7	1260	9	BC016336	BC016336 Homo sapi
29		34.6	14.7	1293	9	AB000905	AB000905 Homo sapi
30		34.6	14.7	89016	9	HS86C11	AL021807 Human DNA
31	c	34.6	14.7	141991	2	HSDJ821D9	AL121960 Homo sapi
32		34.2	14.6	177806	9	AC093106	AC093106 Homo sapi
33	c	34.2	14.6	197246	2	AC118774	AC118774 Rattus no
34		34.2	14.6	269131	2	AC135469	AC135469 Mus muscu
35		34.2	14.6	281804	2	AC134869	AC134869 Mus muscu
36		33.8	14.4	76471	9	AC093743	AC093743 Homo sapi
37	c	33.8	14.4	252473	2	AC093972	AC093972 Rattus no
38		33.8	14.4	271368	2	AC134642	AC134642 Rattus no
39		33.6	14.3	4745	6	BD174728	BD174728 Novel iso
40		33.4	14.2	172105	2	AC021197	AC021197 Homo sapi
41		33.4	14.2	196512	9	AC020911	AC020911 Homo sapi
42		33.2	14.1	175945	2	AC022147	AC022147 Homo sapi
43	c	33	14.0	2544	9	BC006344	BC006344 Homo sapi
44	c	33	14.0	2732	9	BC001928	BC001928 Homo sapi
45	c	33	14.0	2741	9	BC000425	BC000425 Homo sapi

ALIGNMENTS

RESULT 1
AX526280
LOCUS AX526280 235 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 1 from Patent WO02066071.
ACCESSION AX526280
VERSION AX526280.1 GI:25171090
KEYWORDS .
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1
AUTHORS Mauviel,A.
TITLE Blocking sp1 transcription factor broadly inhibits extracellular matrix gene expression in vitro and in vivo: implications for the treatments of tissue fibrosis

```
  JOURNAL   Patent: WO 02066071-A 1 29-AUG-2002;
            Thomas Jefferson University (US)
FEATURES             Location/Qualifiers
     source          1. .235
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
                     /note="oligonucleotide"
BASE COUNT       60 a     64 c     57 g     54 t
ORIGIN

  Query Match             100.0%;  Score 235;  DB 6;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-64;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235


RESULT 2
HUMTFSP1
LOCUS       HUMTFSP1                2655 bp    mRNA    linear   PRI 14-JAN-1995
DEFINITION  Human transcription factor SP1 mRNA, 3' end.
ACCESSION   J03133
VERSION     J03133.1  GI:339517
KEYWORDS    transcription factor SP1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2655)
  AUTHORS   Kadonaga,J.T., Carner,K.R., Masiarz,F.R. and Tjian,R.
  JOURNAL   Unpublished (1987)
REFERENCE   2  (sites)
  AUTHORS   Kadonaga,J.T., Carner,K.R., Masiarz,F.R. and Tjian,R.
  TITLE     Isolation of cDNA encoding transcription factor Sp1 and functional
            analysis of the DNA binding domain
  JOURNAL   Cell 51 (6), 1079-1090 (1987)
  MEDLINE   88080466
   PUBMED   3319186
COMMENT     Original source text: Human HeLa cell, cDNA to mRNA, (library of
            C.Hauser), clone Sp1-1.
            [2]  sites; binding domains.
            Draft entry and computer-readable sequence for [1] kindly provided
            by J.T.Kadonaga, 17-NOV-1987.
FEATURES             Location/Qualifiers
     source          1. .2655
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="12q"
     gene            1. .2655
                     /gene="SP1"
     CDS             <1. .2091
                     /gene="SP1"
                     /note="transcription factor Sp-1"
                     /codon_start=1
                     /protein_id="AAA61154.1"
                     /db_xref="GI:339518"
                     /db_xref="GDB:G00-127-453"
                     /translation="TGELDLTATQLSQGANGWQIISSSSGATPTSKEQSGSSTNGSNG
                     SESSKNRTVSGGQYVVAAAPNLQNQQVLTGLPGVMPNIQYQVIPQFQTVDGQQLQFAA
                     TGAQVQQDGSGQIQIIPGANQQIITNRGSGGNIIAAMPNLLQQAVPLQGLANNVLSGQ
                     TQYVTNVPVALNGNITLLPVNSVSAATLTPSSQAVTISSSGSQESGSQPVTSGTTISS
                     ASLVSSQASSSSFFTNANSYSTTTTSNMGIMNFTTSGSSGTNSQGQTPQRVSGLQGS
                     DALNIQQNQTSGGSLQAGQQKEGEQNQQTQQQQILIQPQLVQGGQALQALQAAPLSGQ
                     TFTTQAISQETLQNLQLQAVPNSGPIIIRTPTVGPNGQVSWQTLQLQNLQVQNPQAQT
                     ITLAPMQGVSLGQTSSSNTTLTPIASAASIPAGTVTVNAAQLSSMPGLQTINLSALGT
                     SGIQVHPIQGLPLAIANAPGDHGAQLGLHGAGGDGIHDDTAGGEEGENSPDAQPQAGR
                     RTRREACTCPYCKDSEGRGSGDPGKKKQHICHIQGCGKVYGKTSHLRAHLRWHTGERP
                     FMCTWSYCGKRFTRSDELQRHKRTHTGEKKFACPECPKRFMRSDHLSKHIKTHQNKKG
                     GPGVALSVGTLPLDSGAGSEGSGTATPSALITTNMVAMEAICPEGIARLANSGINVMQ
                     VADLQSINISGNGF"
BASE COUNT      717 a    726 c    630 g    582 t
ORIGIN      540 bp upstream of HincII site.

  Query Match             100.0%;  Score 235;  DB 9;  Length 2655;
  Best Local Similarity   100.0%;  Pred. No. 1e-63;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 106

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 166

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 226

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 281


RESULT 3
BC043224
LOCUS       BC043224                2806 bp    mRNA    linear   PRI 31-JAN-2003
DEFINITION  Homo sapiens, similar to Transcription factor Sp1, clone
            IMAGE:5295465, mRNA, partial cds.
ACCESSION   BC043224
VERSION     BC043224.1  GI:28175027
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2806)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
```

George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 76 Row: i Column: 12.

```
FEATURES             Location/Qualifiers
     source          1. .2806
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5295465"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             <1. .2397
                     /codon_start=1
                     /product="similar to Transcription factor Sp1"
                     /protein_id="AAH43224.1"
                     /db_xref="GI:28175028"
                     /translation="SGGPARANRLPGPPSDQDHSMDEMTAVVKIEKGVGGNNGGNGNG
                     GGAFSQARSSSTGSSSSTGGGGQESQPSPLALLAATCSRIESPNENSNNSQGPSQSGG
                     TGELDLTATQLSQGANGWQIISSSSGATPTSKEQSGSSTNGSNGSESSKNRTVSGGQY
                     VVAAAPNLQNQQVLTGLPGVMPNIQYQVIPQFQTVDGQQLQFAATGAQVQQDGSGQIQ
                     IIPGANQQIITNRGSGGNIIAAMPNLLQQAVPLQGLANNVLSGQTQYVTNVPVALNGN
                     ITLLPVNSVSAATLTPSSQAVTISSSGSQESGSQPVTSGTTISSASLVSSQASSSSFF
                     TNANSYSTTTTTSNMGIMNFTTSGSSGTNSQGQTPQRVSGLQGSDALNIQQNQTSGGS
                     LQAGQQKEGEQNQQTQQQQILIQPQLVQGGQALQALQAAPLSGQTFTTQAISQETLQN
                     LQLQAVPNSGPIIIRTPTVGPNGQVSWQTLQLQNLQVQNPQAQTITLAPMQGVSLGQT
                     SSSNTTLTPIASAASIPAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHPIQGLPLA
                     IANAPGDHGAQLGLHGAGGDGIHDDTAGGEEGENSPDAQPQAGRRTRREACTCPYCKD
                     SEGRGSGDPGKKKQHICHIQGCGKVYGKTSHLRAHLRWHTGERPFMCTWSYCGKRFTR
                     SDELQRHKRTHTGEKKFACPECPKRFMRSDHLSKHIKTHQNKKGGPGVALSVGTLPLD
                     SGAGSEGSGTATPSALITTNMVAMEAICPEGIARLANSGINVMQVADLQSINISGNGF
                     "
BASE COUNT      756 a    779 c    698 g    573 t
ORIGIN
```

```
  Query Match             100.0%;  Score 235;  DB 9;  Length 2806;
  Best Local Similarity   100.0%;  Pred. No. 1e-63;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 412

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 472

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        473 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 532

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 587
```

```
RESULT 4
AF252284
LOCUS       AF252284               2913 bp    mRNA    linear   PRI 04-MAY-2000
DEFINITION  Homo sapiens transcription specificity factor Sp1 mRNA, partial
            cds.
ACCESSION   AF252284
VERSION     AF252284.1  GI:7689255
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2913)
  AUTHORS   Haggart,M.H. and Ladurner,A.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2000) Molecular and Cell Biology, University of
            California at Berkeley, 401 Barker Hall, No. 3204, Berkeley, CA
            94720-3204, USA
FEATURES             Location/Qualifiers
     source          1. .2913
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="HeLa"
                     /tissue_type="cervical carcinoma"
     CDS             <1. .2349
                     /note="transcriptional activator protein; ubiquitous
                     enhancer binding protein; binds GC-boxes"
                     /codon_start=1
                     /product="transcription specificity factor Sp1"
                     /protein_id="AAF67726.1"
                     /db_xref="GI:7689256"
                     /translation="QDHSMDEMTAVVKIEKGVGGNNGGNGNGGGAFSQARSSSTGSSS
                     STGGGGQESQPSPLALLAATCSRIESPNENSNNSQGPSQSGGTGELDLTATQLSQGAN
                     GWQIISSSSGATPTSKEQSGSSTNGSNGSESSKNRTVSGGQYVVAAAPNLQNQQVLTG
                     LPGVMPNIQYQVIPQFQTVDGQQLQFAATGAQVQQDGSGQIQIIPGANQQIITNRGSG
                     GNIIAAMPNLLQQAVPLQGLANNVLSGQTQYVTNVPVALNGNITLLPVNSVSAATLTP
                     SSQAVTISSSGSQESGSQPVTSGTTISSASLVSSQASSSSFFTNANSYSTTTTTSNMG
                     IMNFTTSGSSGTNSQGQTPQRVSGLQGSDALNIQQNQTSGGSLQAGQQKEGEQNQQTQ
                     QQQILIQPQLVQGGQALQALQAAPLSGQTFTTQAISQETLQNLQLQAVPNSGPIIIRT
                     PTVGPNGQVSWQTLQLQNLQVQNPQAQTITLAPMQGVSLGQTSSSNTTLTPIASAASI
                     PAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHPIQGLPLAIANAPGDHGAQLGLHG
                     AGGDGIHDDTAGGEEGENSPDAQPQAGRRTRREACTCPYCKDSEGRGSGDPGKKKQHI
                     CHIQGCGKVYGKTSHLRAHLRWHTGERPFMCTWSYCGKRFTRSDELQRHKRTHTGEKK
                     FACPECPKRFMRSDHLSKHIKTHQNKKGGPGVALSVGTLPLDSGAGSEGSGTATPSAL
                     ITTNMVAMEAICPEGIARLANSGINVMQVADLQSINISGNGF"
BASE COUNT      790 a    783 c    712 g    628 t
ORIGIN
```

```
  Query Match             100.0%;  Score 235;  DB 9;  Length 2913;
  Best Local Similarity   100.0%;  Pred. No. 1e-63;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 364

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 424

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 484

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 539
```

```
RESULT 5
AB039286
LOCUS       AB039286               4300 bp    DNA     linear   PRI 01-DEC-2000
DEFINITION  Homo sapiens gene for transcription factor Sp1, partial cds.
ACCESSION   AB039286
VERSION     AB039286.1  GI:10119775
KEYWORDS    transcription factor Sp1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Takahara,T., Kanazu,S.I., Yanagisawa,S. and Akanuma,H.
  TITLE     Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of
            homotypic trans-splicing
  JOURNAL   J. Biol. Chem. 275 (48), 38067-38072 (2000)
  MEDLINE   20545561
```

```
   PUBMED   10973950
REFERENCE   2  (bases 1 to 4300)
  AUTHORS   Takahara,T., Yanagisawa,S. and Akanuma,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Hiroshi Akanuma, University of Tokyo,
            Graduate School of Arts & Sciences; Komaba 3-8-1, Meguro, Tokyo
            153-8902, Japan (E-mail:cakanum@m.ecc.u-tokyo.ac.jp,
            Tel:+81-3-5454-4392(ex.44392))
FEATURES             Location/Qualifiers
     source          1. .4300
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="lambdaSp1E1"
                     /cell_line="HepG2 cell"
     gene            join(438. .444,1805. .1959,2256. .3768)
                     /gene="Sp1"
     CDS             join(438. .444,1805. .1959,2256. .>3768)
                     /gene="Sp1"
                     /codon_start=1
                     /product="transcription factor Sp1"
                     /protein_id="BAB13476.1"
                     /db_xref="GI:10119776"
                     /translation="MSDQDHSMDEMTAVVKIEKGVGGNNGGNGNGGGAFSQARSSSTG
                     SSSSTGGGGQESQPSPLALLAATCSRIESPNENSNNSQGPSQSGGTGELDLTATQLSQ
                     GANGWQIISSSSGATPTSKEQSGSSTNGSNGSESSKNRTVSGGQYVVAAAPNLQNQQV
                     LTGLPGVMPNIQYQVIPQFQTVDGQQLQFAATGAQVQQDGSGQIQIIPGANQQIITNR
                     GSGGNIIAAMPNLLQQAVPLQGLANNVLSGQTQYVTNVPVALNGNITLLPVNSVSAAT
                     LTPSSQAVTISSSGSQESGSQPVTSGTTISSASLVSSQASSSSFFTNANSYSTTTTTS
                     NMGIMNFTTSGSSGTNSQGQTPQRVSGLQGSDALNIQQNQTSGGSLQAGQQKEGEQNQ
                     QTQQQQILIQPQLVQGGQALQALQAAPLSGQTFTTQAISQETLQNLQLQAVPNSGPII
                     IRTPTVGPNGQVSWQTLQLQNLQVQNPQAQTITLAPMQGVSLGQTSSSNTTLTPIASA
                     ASIPAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHPIQGLPLAIANAP"
BASE COUNT      937 a   1260 c   1016 g   1087 t
ORIGIN

  Query Match           100.0%;  Score 235;  DB 9;  Length 4300;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-63;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2407 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 2466

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2467 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 2526

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2527 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 2586

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2587 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 2641


RESULT 6
AC068889/c
LOCUS       AC068889               73433 bp    DNA     linear   PRI 01-JAN-2003
DEFINITION  Homo sapiens 12 BAC RP11-774I22 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC068889
VERSION     AC068889.35  GI:27452896
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 73433)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
            Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
            Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
            Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
            Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
            Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
            King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
            Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
            Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X.,
            Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
            Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
            Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
            Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
            Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
            Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 73433)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 73433)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 73433)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 1, 2003 this sequence version replaced gi:23264934.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
```

local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

```
FEATURES             Location/Qualifiers
     source          1. .73433
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-774I22"
     repeat_region   12. .66
                     /rpt_family="(CGGG)n"
     misc_feature    75. .2004
                     /note="overlaps bases 170628. .172558 of clone AC023509"
                     /function="clone overlap"
     repeat_region   505. .799
                     /rpt_family="L1MC5"
     repeat_region   842. .888
                     /rpt_family="(TA)n"
     repeat_region   915. .1149
                     /rpt_family="AluJo"
     repeat_region   complement(1166. .1289)
                     /rpt_family="AluSp/q"
     repeat_region   1308. .1487
                     /rpt_family="L1MC4"
     repeat_region   1428. .1521
                     /rpt_family="L1MC5"
     repeat_region   1608. .1623
                     /rpt_family="AluJo"
     repeat_region   1624. .1923
                     /rpt_family="AluSp"
     repeat_region   1924. .2066
                     /rpt_family="AluJo"
     repeat_region   2100. .2162
                     /rpt_family="L1MC5"
     repeat_region   complement(2163. .2467)
                     /rpt_family="AluSg"
     repeat_region   2468. .2537
                     /rpt_family="L1MC5"
     repeat_region   2538. .2846
                     /rpt_family="AluSc"
     repeat_region   2847. .2962
                     /rpt_family="L1MC5"
     repeat_region   2968. .3267
                     /rpt_family="AluSc"
     repeat_region   3269. .3447
                     /rpt_family="AluJb"
     repeat_region   3457. .3714
                     /rpt_family="AluJb"
     repeat_region   3872. .4141
                     /rpt_family="AluSp"
     repeat_region   complement(4142. .4197)
                     /rpt_family="Alu"
     repeat_region   4538. .4588
                     /rpt_family="MIR"
     repeat_region   complement(4613. .4854)
                     /rpt_family="AluSp"
     repeat_region   4904. .5211
                     /rpt_family="AluSq"
     repeat_region   5220. .5254
                     /rpt_family="AT_rich"
     STS             5615. .5787
                     /standard_name="SHGC-187"
     STS             5624. .5762
                     /standard_name="D3S4196"
     repeat_region   complement(6853. .7147)
                     /rpt_family="AluSg"
     repeat_region   complement(7303. .7515)
                     /rpt_family="AluSg/x"
     repeat_region   7589. .7887
                     /rpt_family="AluY"
     repeat_region   9063. .9327
                     /rpt_family="AluSx"
     repeat_region   9675. .9955
                     /rpt_family="AluSg"
     repeat_region   9956. .9978
                     /rpt_family="(GAAAA)n"
     STS             10267. .10546
                     /standard_name="G64247"
     repeat_region   10966. .11220
                     /rpt_family="L2"
     repeat_region   complement(11221. .11509)
                     /rpt_family="AluSc"
     repeat_region   11524. .11535
                     /rpt_family="AluSg/x"
     repeat_region   11536. .11839
                     /rpt_family="AluSg"
     misc_feature    11548
                     /note="Position below phred 30 quality."
                     /function="low quality"
     repeat_region   11840. .12010
                     /rpt_family="AluSg/x"
```

```
  Query Match             100.0%;  Score 235;  DB 9;  Length 73433;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-63;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69956 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 69897

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69896 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 69837

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69836 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 69777

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69776 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 69722


RESULT 7
RATSP1
LOCUS       RATSP1                  2475 bp    mRNA    linear   ROD 28-SEP-1999
DEFINITION  Rattus norvegicus mRNA for Sp1, complete cds.
ACCESSION   D12768
VERSION     D12768.1  GI:220911
KEYWORDS    Sp1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
```

```
            Rattus.
REFERENCE   1  (bases 1 to 2475)
  AUTHORS   Imataka,H., Sogawa,K., Yasumoto,K., Kikuchi,Y., Sasano,K.,
            Kobayashi,A., Hayami,M. and Fujii-Kuriyama,Y.
  TITLE     Two regulatory proteins that bind to the basic transcription
            element (BTE), a GC box sequence in the promoter region of the rat
            P-4501A1 gene
  JOURNAL   EMBO J. 11 (10), 3663-3671 (1992)
  MEDLINE   93010958
   PUBMED   1356762
REFERENCE   2  (bases 1 to 2475)
  AUTHORS   Fujii-Kuriyama,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1992) Yoshiaki Fujii-Kuriyama, Faculty of
            Science, Tohoku University, Department of Chemistry; Aramaki
            Aoba-ku, Sendai, Miyagi 980-77, Japan (Tel:022-222-1800(ex.3380),
            Fax:022-262-6609)
FEATURES             Location/Qualifiers
     source          1. .2475
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     CDS             49. .2415
                     /codon_start=1
                     /product="Sp1"
                     /protein_id="BAA02235.1"
                     /db_xref="GI:220912"
                     /translation="MSDQDHSMDEVTAVKIEKGVGGNNGGSGNGGGAAFSQTRSSSTG
                     SSSSSGGGGGQESQPSPLALLAATCSRIESPNENENSNNSQGPSQSGGTGELDLTATQ
                     LSQGANGWQIISSSSGATPTSKEQSGNSTNGSNGSESSKNRTVSGGQYVVAATPNLQN
                     QQVLTGLPGVMPNIQYQVIPQFQTVDGQQLQFAATGAQVQQDGSGQIQIIPGANQQII
                     TNRGSGGNIIAAMPNLLQQAVPLQGLANNVLSGQTQYVTNVPVALNGNITLLPVNSVS
                     AATLTPSSQAGTISSSGSQESGSQPVTSGTAISSASLVSSQASSSSFFTNANSYSTTT
                     TTSNMGIMNFTSSGSSGTSSQGQTSQRVGGLQGSDSLNIQQNQTSGGSLQGSQQKEGE
                     QSQQTQQQQILIQPQLVQGGQALQLQAAPLSGQTFTTQAISQETLQNLQLQAVQNSG
                     PIIIRTPTVGPNGQVSWQTLQLQNLQVQNPQAQTITLAPMQGVSLGQTSSSNTTLTPI
                     ASAASIPAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHQLPGLPLAIANTPGDHGA
                     QLGLHGPGGDGIHDETAGGEEGENSPDPQPQAGRRTRREACTCPYCKDSEGRGSGDPG
                     KKKQHICHIQGCGKVYGKTSHLRAHLRWHTGERPFMCNWSYCGKRFTRSDELQRHKRT
                     HTGEKKFACPECPKRFMRSDHLSKHIKTHQNKKGGPGVALSVGTLPLDSGAGSEGSGT
                     ATPSALITTNMVAMEAICPEGIARLANSGINVMQVTELQSINISGNGF"
BASE COUNT      668 a    683 c    626 g    498 t
ORIGIN

  Query Match           94.6%;  Score 222.2;  DB 10;  Length 2475;
  Best Local Similarity  96.6%;  Pred. No. 1.2e-59;
  Matches  227;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 430

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||| |||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db        431 GCAACAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAACCGCACAGTCTCTGGTG 490

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||||| |||||||||||||||||||||||||||||| ||||
Db        491 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTAC 550

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| ||||||||||||||||||||||||||||||||||||| |||||
Db        551 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 605


RESULT 8
AB077988
LOCUS       AB077988               4092 bp    DNA     linear   ROD 23-OCT-2002
DEFINITION  Rattus norvegicus gene for transcription factor Sp1, partial cds.
ACCESSION   AB077988
VERSION     AB077988.1  GI:21832059
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Takahara,T., Kasahara,D., Mori,D., Yanagisawa,S. and Akanuma,H.
  TITLE     The trans-spliced variants of Sp1 mRNA in rat
  JOURNAL   Biochem. Biophys. Res. Commun. 298 (1), 156-162 (2002)
  MEDLINE   22266831
   PUBMED   12379234
REFERENCE   2  (bases 1 to 4092)
  AUTHORS   Mori,D., Takahara,T., Yanagisawa,S. and Akanuma,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2002) Hiroshi Akanuma, University of Tokyo,
            Graduate School of Arts & Sciences; Komaba, 3-8-1, Meguro, Tokyo
            153-8902, Japan (E-mail:cakanum@m.ecc.u-tokyo.ac.jp,
            Tel:81-3-5454-4392(ex.44392))
FEATURES             Location/Qualifiers
     source          1. .4092
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /tissue_type="liver"
                     /clone_lib="rat genomic library (Clontech, RL1022j))"
     gene            join(324. .330,1673. .1830,2112. .3624)
                     /gene="Sp1"
     CDS             join(324. .330,1673. .1830,2112. .>3624)
                     /gene="Sp1"
                     /codon_start=1
                     /product="transcription factor Sp1"
                     /protein_id="BAC05486.1"
                     /db_xref="GI:21832060"
                     /translation="MSDQDHSMDEVTAVKIEKGVGGNNGGSGNGGGAAFSQTRSSSTG
                     SSSSSGGGGGQESQPSPLALLAATCSRIESPNENSNNSQGPSQSGGTGELDLTATQLS
                     QGANGWQIISSSSGATPTSKEQSGNSTNGSNGSESSKNRTVSGGQYVVAATPNLQNQQ
                     VLTGLPGVMPNIQYQVIPQFQTVDGQQLQFAATGAQVQQDGSGQIQIIPGANQQIITN
                     RGSGGNIIAAMPNLLQQAVPLQGLANNVLSGQTQYVTNVPVALNGNITLLPVNSVSAA
                     TLTPSSQAGTISSSGSQESGSQPVTSGTAISSASLVSSQASSSSFFTNANSYSTTTTT
                     SNMGIMNFTSSGSSGTSSQGQTSQRVGGLQGSDSLNIQQNQTSGGSLQGSQQKEGEQS
                     QQTQQQQILIQPQLVQGGQALQALQAAPLSGQTFTTQAISQETLQNLQLQAVQNSGPI
                     IIRTPTVGPNGQVSWQTLQLQNLQVQNPQAQTITLAPMQGVSLGQTSSSNTTLTPIAS
                     AASIPAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHQLPGLPLAIANTP"
BASE COUNT      943 a   1155 c    979 g   1015 t
ORIGIN

  Query Match           94.6%;  Score 222.2;  DB 10;  Length 4092;
  Best Local Similarity  96.6%;  Pred. No. 1.3e-59;
  Matches  227;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2263 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 2322

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||| |||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       2323 GCAACAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAACCGCACAGTCTCTGGTG 2382

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||||| |||||||||||||||||||||||||||||| ||||
Db       2383 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTAC 2442

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| ||||||||||||||||||||||||||||||||||||| |||||
Db       2443 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 2497


RESULT 9
AC109743
LOCUS       AC109743               206520 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-65D10, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC109743
```

```
VERSION     AC109743.5  GI:25072885
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 206520)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 206520)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 206520)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23195412.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GQNJ
                Center clone name: CH230-65D10
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 168784 bases at least Q40
                Consensus quality: 173457 bases at least Q30
                Consensus quality: 175960 bases at least Q20
                Estimated insert size: 178632; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    76730: contig of 76730 bp in length
            *   76731    76830: gap of unknown length
            *   76831   204784: contig of 127954 bp in length
            *  204785   204884: gap of unknown length
            *  204885   206520: contig of 1636 bp in length.
FEATURES             Location/Qualifiers
     source          1. .206520
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-65D10"
     misc_feature    101270. .102941
                     /note="wgs_contig"
     misc_feature    168407. .170678
                     /note="wgs_contig"
BASE COUNT    49558 a  37786 c  38536 g  52117 t  28523 others
ORIGIN

  Query Match             94.6%;  Score 222.2;  DB 2;  Length 206520;
  Best Local Similarity   96.6%;  Pred. No. 1.9e-59;
  Matches  227;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14467 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 14526

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      14527 GCAACAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAACCGCACAGTCTCTGGTG 14586

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||||| | ||||||||||||||||||||||||||| ||||
Db      14587 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTAC 14646

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| ||||||||||||||||||||||||||||||||||||| |||||
```

```
Db      14647 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 14701


RESULT 10
AC097309
LOCUS       AC097309               254686 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-97M10, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC097309
VERSION     AC097309.6  GI:30521315
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 254686)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 254686)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 254686)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23268851.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GIGO
                Center clone name: CH230-97M10
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 240288 bases at least Q40
                Consensus quality: 242808 bases at least Q30
                Consensus quality: 244806 bases at least Q20
                Estimated insert size: 254704; sum-of-contigs estimation
                Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   253477: contig of 253477 bp in length
            *  253478   253577: gap of unknown length
            *  253578   254686: contig of 1109 bp in length.
FEATURES             Location/Qualifiers
     source          1. .254686
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-97M10"
     misc_feature    1. .3186
                     /note="wgs_contig"
     misc_feature    251549. .252396
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH283956"
BASE COUNT    63545 a  58875 c  59227 g  63989 t   9050 others
ORIGIN

  Query Match             94.6%;  Score 222.2;  DB 2;  Length 254686;
  Best Local Similarity   96.6%;  Pred. No. 1.9e-59;
  Matches  227;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212410 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 212469
```

```
Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     212470 GCAACAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAACCGCACAGTCTCTGGTG 212529

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||||  | ||||||||||||||||||||||||||||||| ||||
Db     212530 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTAC 212589

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| |||||||||||||||||||||||||||||||||||||| |||||
Db     212590 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 212644


RESULT 11
AC021103
LOCUS       AC021103               166697 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome UL clone RP11-147A18, WORKING DRAFT
            SEQUENCE, 43 unordered pieces.
ACCESSION   AC021103
VERSION     AC021103.7  GI:8099088
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166697)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166697)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 26, 2000 this sequence version replaced gi:8018258.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0147A18
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 66%
            Sequencing vector: plasmid; 34%
            Chemistry: Dye-primer ET; 66% of reads
            Chemistry: Dye-terminator Big Dye; 34% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 151987 bases at least Q40
            Consensus quality: 156240 bases at least Q30
            Consensus quality: 158337 bases at least Q20
            Insert size: 184000; agarose-fp
            Insert size: 162497; sum-of-contigs
            Quality coverage: 3.88 in Q20 bases; agarose-fp
            Quality coverage: 4.39 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 43 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1466: contig of 1466 bp in length
            *    1467     1566: gap of unknown length
            *    1567     2896: contig of 1330 bp in length
            *    2897     2996: gap of unknown length
            *    2997     4016: contig of 1020 bp in length
            *    4017     4116: gap of unknown length
            *    4117     5406: contig of 1290 bp in length
            *    5407     5506: gap of unknown length
            *    5507     6642: contig of 1136 bp in length
            *    6643     6742: gap of unknown length
            *    6743     8257: contig of 1515 bp in length
            *    8258     8357: gap of unknown length
            *    8358    10742: contig of 2385 bp in length
            *   10743    10842: gap of unknown length
            *   10843    12190: contig of 1348 bp in length
            *   12191    12290: gap of unknown length
            *   12291    13569: contig of 1279 bp in length
            *   13570    13669: gap of unknown length
            *   13670    15855: contig of 2186 bp in length
            *   15856    15955: gap of unknown length
            *   15956    17977: contig of 2022 bp in length
            *   17978    18077: gap of unknown length
            *   18078    20177: contig of 2100 bp in length
            *   20178    20277: gap of unknown length
            *   20278    22120: contig of 1843 bp in length
            *   22121    22220: gap of unknown length
            *   22221    24582: contig of 2362 bp in length
            *   24583    24682: gap of unknown length
            *   24683    27320: contig of 2638 bp in length
            *   27321    27420: gap of unknown length
            *   27421    29908: contig of 2488 bp in length
            *   29909    30008: gap of unknown length
            *   30009    32534: contig of 2526 bp in length
            *   32535    32634: gap of unknown length
            *   32635    35493: contig of 2859 bp in length
            *   35494    35593: gap of unknown length
            *   35594    37551: contig of 1958 bp in length
            *   37552    37651: gap of unknown length
            *   37652    41357: contig of 3706 bp in length
            *   41358    41457: gap of unknown length
            *   41458    43778: contig of 2321 bp in length
            *   43779    43878: gap of unknown length
            *   43879    47504: contig of 3626 bp in length
            *   47505    47604: gap of unknown length
            *   47605    50117: contig of 2513 bp in length
            *   50118    50217: gap of unknown length
            *   50218    53652: contig of 3435 bp in length
            *   53653    53752: gap of unknown length
            *   53753    56657: contig of 2905 bp in length
            *   56658    56757: gap of unknown length
            *   56758    60073: contig of 3316 bp in length
            *   60074    60173: gap of unknown length
            *   60174    64263: contig of 4090 bp in length
            *   64264    64363: gap of unknown length
            *   64364    68859: contig of 4496 bp in length
            *   68860    68959: gap of unknown length
            *   68960    72754: contig of 3795 bp in length
            *   72755    72854: gap of unknown length
            *   72855    77167: contig of 4313 bp in length
            *   77168    77267: gap of unknown length
            *   77268    81554: contig of 4287 bp in length
            *   81555    81654: gap of unknown length
            *   81655    84716: contig of 3062 bp in length
            *   84717    84816: gap of unknown length
            *   84817    89925: contig of 5109 bp in length
            *   89926    90025: gap of unknown length
            *   90026    94331: contig of 4306 bp in length
            *   94332    94431: gap of unknown length
            *   94432    98564: contig of 4133 bp in length
            *   98565    98664: gap of unknown length
            *   98665   105959: contig of 7295 bp in length
            *  105960   106059: gap of unknown length
            *  106060   112344: contig of 6285 bp in length
            *  112345   112444: gap of unknown length
            *  112445   119806: contig of 7362 bp in length
            *  119807   119906: gap of unknown length
            *  119907   127259: contig of 7353 bp in length
            *  127260   127359: gap of unknown length
            *  127360   136297: contig of 8938 bp in length
            *  136298   136397: gap of unknown length
```

```
           *   136398   144884: contig of 8487 bp in length
           *   144885   144984: gap of unknown length
           *   144985   154492: contig of 9508 bp in length
           *   154493   154592: gap of unknown length
           *   154593   166697: contig of 12105 bp in length.
FEATURES             Location/Qualifiers
     source          1. .166697
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="UL"
                     /clone="RP11-147A18"
BASE COUNT    41526 a  40406 c  40099 g  40461 t   4205 others
ORIGIN

  Query Match           94.0%;  Score 221;  DB 2;  Length 166697;
  Best Local Similarity   99.1%;  Pred. No. 4.4e-59;
  Matches  233;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    1;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145287 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 145346

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145347 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 145406

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db     145407 GGCAGTATGTTGTG--TGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 145464

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145465 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 145519

.
RESULT 12
AF022363
LOCUS       AF022363                4128 bp    mRNA    linear   ROD 31-MAR-1998
DEFINITION  Mus musculus transcription factor SP1 (Sp1) mRNA, complete cds.
ACCESSION   AF022363
VERSION     AF022363.1  GI:2997600
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4128)
  AUTHORS   Park,E.J., Kim,J.H., Kim,C.G., Park,S.D. and Hong,S.S.
  TITLE     Isolation of cDNA encoding transcription factor sp1 containing two
            glutamine-rich transactivation domain
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4128)
  AUTHORS   Park,E.J., Kim,J.H., Kim,C.G., Park,S.D. and Hong,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1997) Molecular Biology, IMBG, Shilim-dong,
            Kwanak-gu, Seoul 151-742, Korea
FEATURES             Location/Qualifiers
     source          1. .4128
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="thymus"
     gene            1. .4128
                     /gene="Sp1"
     CDS             89. .2443
                     /gene="Sp1"
                     /codon_start=1
                     /product="transcription factor SP1"
                     /protein_id="AAC08527.1"
                     /db_xref="GI:2997601"
                     /translation="MSDQDHSMDEVTAVVKIEKDVGGNNGGSGNGGGAAFSQTRSSST
                     GSSSSSGGGGGQESQPSPLALLAATCSRIESPNENSNNSQGPSQSGGTGELDLTATQL
                     SQGANGWQIISSSSGATPTSKEQSGNSTNGSNGSESSKNRTVSGGQYVVAATPNLQNQ
                     QVLTGLPGVMPNIQYQVIPQFQTVDGQQLQFAATGAQVQQDGSGQIQIIPGANQQIIP
                     NEGSGGNIIAAMPNLLQQAVPLQGLANNVLSGQTQYVTNVPVALNGNITLLPVNSVSA
                     ATLTPSSQAGTISSSGSQESSSQPVTSGTAISSASLVSSQASSSSFFTNANSYSTTTT
                     TSNMGIMNFTSSGSSGTSSQGQTPQRVGGLQGSDSLNIQQNQTSGGSLQGSQQKEGEQ
                     SQQTQQQQILIQPQLVQGGQALQALQAAPLSGQTFTTQAISQETLQNLQLQAVQNSGP
                     IIIRTPTVGPNGQVSWQTLQLQNLQVQNPQAQTITLAPMQGVSLGQTSSSNTTLTPIA
                     SAASIPAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHQLPGLPLAIANTPGDHGTQ
                     LGLHGSGGDGIHDETAGGEGENSSDLQPQAGRRTRREACTCPYCKDSEGRASGDPGKK
                     KQHICHIQGCGKVYGKTSHLRAHLRWHTGERPFMCNWSYCGKRFTRSDELQRHKRTHT
                     GEKKFACPECPKRFMRSDHLSKHIKTHQNKKGGPGVALSVGTLPLDSGAGSEGTATPS
                     ALITTNMVAMEAICPEGIARLANSGINVMQVTELQSINISGNGF"
BASE COUNT     1177 a   1009 c    922 g   1020 t
ORIGIN

  Query Match           93.9%;  Score 220.6;  DB 10;  Length 4128;
  Best Local Similarity   96.2%;  Pred. No. 4.2e-59;
  Matches  226;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 467

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||| |||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        468 GCAACAGTACCAATGGCAGCAATGGCAGCGAGTCTTCCAAGAACCGCACAGTCTCTGGTG 527

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||  | |||||||||||||||||||||||||||| ||||
Db        528 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTAC 587

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| |||||||||||||||||||||||||||||||||||||| |||||
Db        588 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 642


RESULT 13
G92513
LOCUS       G92513                  586 bp    DNA     linear   STS 06-SEP-2002
DEFINITION  S208P6120RE3.T0 129S1/SvImJ Mus musculus STS genomic, sequence
            tagged site.
ACCESSION   G92513
VERSION     G92513.1  GI:22743269
KEYWORDS    STS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Wade,C.
  TITLE     Polymorphism Structure in the Mouse
  JOURNAL   Unpublished (2002)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kerstli@genome.wi.mit.edu
            Primer A: None
            Primer B: None
            STS size: 586
            Protocol:
            WGS-discovery: Paired-end low-coverage whole genome shotgun reads
            were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
            reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
            detection was carried out by SSAHA-SNP. 225,000 reads were
            annotated
            as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
            and the strain from which the particular read came. The validation
            rate for these SNPs was estimated at approximately 98%.
FEATURES             Location/Qualifiers
```

```
     source          1. .586
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="129S1/SvImJ"
                     /db_xref="taxon:10090"
                     /map="+ 15 22-551 103388565-103389094"
                     /clone_lib="129S1/SvImJ"
     STS             <1. .>586
BASE COUNT      141 a    159 c    138 g    148 t
ORIGIN

  Query Match             82.0%;  Score 192.6;  DB 11;  Length 586;
  Best Local Similarity   92.3%;  Pred. No. 3e-50;
  Matches  217;  Conservative    0;  Mismatches    9;  Indels    9;  Gaps    1;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 392

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||| ||||||||||||||||          |||||||||||||| ||||||||||||||||
Db        393 GCAACAGTACCAATGGCAGC---------GAGTCTTCCAAGAACCGCACAGTCTCTGGTG 443

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||||  | ||||||||||||||||||||||||||||| || ||
Db        444 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTCC 503

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| ||||||||||||||||||||||||||||||||||||| |||||
Db        504 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 558


RESULT 14
AF062566
LOCUS       AF062566               3741 bp    mRNA    linear   ROD 22-JAN-2001
DEFINITION  Mus musculus transcription factor Sp1 mRNA, complete cds.
ACCESSION   AF062566
VERSION     AF062566.1  GI:3135322
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3741)
  AUTHORS   Yajima,S., Lee,S.H., Minowa,T. and Mouradian,M.M.
  TITLE     Sp family transcription factors regulate expression of rat D2
            dopamine receptor gene
  JOURNAL   DNA Cell Biol. 17 (5), 471-479 (1998)
  MEDLINE   98290554
   PUBMED   9628590
REFERENCE   2  (bases 1 to 3741)
  AUTHORS   Yajima,S., Lee,S.-H., Minowa,T. and Mouradian,M.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-1998) GPU/ETB, NINDS/NIH, 10 Center Dr MSC1406,
            Bethesda, MD 20892-1406, USA
FEATURES             Location/Qualifiers
     source          1. .3741
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /cell_line="NB41A3"
                     /tissue_type="neuroblastoma"
     CDS             37. .2382
                     /codon_start=1
                     /product="transcription factor Sp1"
                     /protein_id="AAC16484.1"
                     /db_xref="GI:3135323"
                     /translation="MSDQDHSMDEVTAVVKIEKDVGGNNGGSGNGGGAAFSQTRSSST
                     GSSSSSGGGGGQESQPSPLALLAATCSRIESPNENSNNSQGPSQSGGTGELDLTAAQL
                     SQGANGWQIISSSSGATPTSKEQSGNSTNGSESSKNRTVSGGQYVVAATPNLQNQQVL
                     TGLPGVMPNIQYQVIPQFQTVDGQQLQFAATGAQVQQDGSGQIQIIPGANQQIIPNRG
                     SGGNIIAAMPNLLQQAVPLQGLANNVLSGQTQYVTNVPVALNGNITLLPVNSVSAATL
                     TPSSQAGTISSSGSQESSSQPVTSGTAISSASLVSSQASSSSFFTNANSYSTTTTTSN
                     MGIMNFTSSGSSGTSSQGQTPQRVGGLQGSDSLNIQQNQTSGGSLQGSQQKEGEQSQQ
                     TQQQQILIQPQLVQGGQALQALQAAPLSGQTFTTQAISQETLQNLQLQAVQNSGPIII
                     RTPTVGPNVQVSWQTLQLQNLQVQNPQAQTITLAPMQGVSLGQTSSSNTTLTPIASAA
                     SIPAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHQLPGLPLAIANTPGDHGTQLGL
                     HGSGGDGIHDETAGGEGENSSDLQPQAGRRTRREACTCPYCKDSEGRASGDPGKKKQH
                     ICHIQGCGKVYGKTSHLRAHLRWHTGERPFMCNWSYCGKRFTRSDELQRHKRTHTGEK
                     KFACPECPKRFMRSDHLSKHIKTHQNKKGGPGVALSVGTLPLDSGAGSEGTATPSALI
                     TTNMVAMEAICPEGIARLANSGINVMQVTELQSINISGNGF"
BASE COUNT     1015 a    928 c    849 g    949 t
ORIGIN

  Query Match             82.0%;  Score 192.6;  DB 10;  Length 3741;
  Best Local Similarity   92.3%;  Pred. No. 3.6e-50;
  Matches  217;  Conservative    0;  Mismatches    9;  Indels    9;  Gaps    1;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 415

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||| ||||||||||||||||          |||||||||||||| ||||||||||||||||
Db        416 GCAACAGTACCAATGGCAGC---------GAGTCTTCCAAGAACCGCACAGTCTCTGGTG 466

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||||  | ||||||||||||||||||||||||||||| || ||
Db        467 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTCC 526

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| ||||||||||||||||||||||||||||||||||||| |||||
Db        527 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 581


RESULT 15
AC055703/c
LOCUS       AC055703               138860 bp    DNA     linear   HTG 15-MAY-2002
DEFINITION  Mus musculus clone RP23-399N14 strain C57BL6/J, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC055703
VERSION     AC055703.9  GI:20279384
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 138860)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE     High Throughput Mouse Sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 138860)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2000) Department of Molecular Genetics, Albert
            Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
            Bronx, NY 10461, USA
REFERENCE   3  (bases 1 to 138860)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
COMMENT     On Apr 24, 2002 this sequence version replaced gi:18151001.
            ----------------Genome Center
            Center:  Harvard Partners Genome Center
            Center Code:  HPGC
            Web site: http://www.hpcgg.org/Sequence/mouse.html
            Contact: hpgc@mendel.mgh.harvard.edu
            ----------------Summary Statistics
```

```
Center project name:  AAS
Sequencing vector:  pUC18; L08752
Chemistry:  Dye-terminator Big Dye; 100%
*Consensus quality:  126666 at least Q20
*Consensus quality:  122952 at least Q30
*Consensus quality:  116499 at least Q40
Estimated insert size:   agarose-FP - N/A
**Estimated insert size: 138140 - sum-of-contigs
Quality coverage:  agarose-FP - N/A
Quality coverage:  9.1 x in Q20 bases; sum-of-contigs estimation
----------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1    10740: contig of 10740 bp in length
*   10741    10760: gap of unknown length
*   10761    18887: contig of 8127 bp in length
*   18888    18907: gap of unknown length
*   18908    26674: contig of 7767 bp in length
*   26675    26694: gap of unknown length
*   26695    31485: contig of 4791 bp in length
*   31486    31505: gap of unknown length
*   31506    36220: contig of 4715 bp in length
*   36221    36240: gap of unknown length
*   36241    43190: contig of 6950 bp in length
*   43191    43210: gap of unknown length
*   43211    47455: contig of 4245 bp in length
*   47456    47475: gap of unknown length
*   47476    52355: contig of 4880 bp in length
*   52356    52375: gap of unknown length
*   52376    57154: contig of 4779 bp in length
*   57155    57174: gap of unknown length
*   57175    63909: contig of 6735 bp in length
*   63910    63929: gap of unknown length
*   63930    70719: contig of 6790 bp in length
*   70720    70739: gap of unknown length
*   70740    73652: contig of 2913 bp in length
*   73653    73672: gap of unknown length
*   73673    79581: contig of 5909 bp in length
*   79582    79601: gap of unknown length
*   79602    83557: contig of 3956 bp in length
*   83558    83577: gap of unknown length
*   83578    89230: contig of 5653 bp in length
*   89231    89250: gap of unknown length
*   89251    93443: contig of 4193 bp in length
*   93444    93463: gap of unknown length
*   93464    97169: contig of 3706 bp in length
*   97170    97189: gap of unknown length
*   97190   102251: contig of 5062 bp in length
*  102252   102271: gap of unknown length
*  102272   107091: contig of 4820 bp in length
*  107092   107111: gap of unknown length
*  107112   110480: contig of 3369 bp in length
*  110481   110500: gap of unknown length
*  110501   113418: contig of 2918 bp in length
*  113419   113438: gap of unknown length
*  113439   116668: contig of 3230 bp in length
*  116669   116688: gap of unknown length
*  116689   118403: contig of 1715 bp in length
*  118404   118423: gap of unknown length
*  118424   120544: contig of 2121 bp in length
*  120545   120564: gap of unknown length
*  120565   121993: contig of 1429 bp in length
*  121994   122013: gap of unknown length
*  122014   122930: contig of 917 bp in length
*  122931   122950: gap of unknown length
*  122951   124509: contig of 1559 bp in length
*  124510   124529: gap of unknown length
*  124530   126851: contig of 2322 bp in length
*  126852   126871: gap of unknown length
*  126872   128496: contig of 1625 bp in length
*  128497   128516: gap of unknown length
*  128517   129949: contig of 1433 bp in length
*  129950   129969: gap of unknown length
*  129970   132003: contig of 2034 bp in length
*  132004   132023: gap of unknown length
*  132024   133605: contig of 1582 bp in length
*  133606   133625: gap of unknown length
*  133626   134100: contig of 475 bp in length
*  134101   134120: gap of unknown length
*  134121   135517: contig of 1397 bp in length
*  135518   135537: gap of unknown length
*  135538   136468: contig of 931 bp in length
*  136469   136488: gap of unknown length
*  136489   137236: contig of 748 bp in length
*  137237   137256: gap of unknown length
*  137257   138860: contig of 1604 bp in length.
FEATURES             Location/Qualifiers
     source          1. .138860
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /clone="RP23-399N14"
                     /sex="male"
     misc_feature    1. .10740
                     /note="assembly_name:Contig270"
     misc_feature    10761. .18887
                     /note="assembly_name:Contig269"
     misc_feature    18908. .26674
                     /note="assembly_name:Contig268"
     misc_feature    26695. .31485
                     /note="assembly_name:Contig267"
     misc_feature    31506. .36220
                     /note="assembly_name:Contig266"
     misc_feature    36241. .43190
                     /note="assembly_name:Contig265"
     misc_feature    43211. .47455
                     /note="assembly_name:Contig264"
     misc_feature    47476. .52355
                     /note="assembly_name:Contig263"
     misc_feature    52376. .57154
                     /note="assembly_name:Contig262"
     misc_feature    57175. .63909
                     /note="assembly_name:Contig261"
     misc_feature    63930. .70719
                     /note="assembly_name:Contig260"
     misc_feature    70740. .73652
                     /note="assembly_name:Contig259"
     misc_feature    73673. .79581
                     /note="assembly_name:Contig258"
     misc_feature    79602. .83557
                     /note="assembly_name:Contig257"
     misc_feature    83578. .89230
                     /note="assembly_name:Contig256"
     misc_feature    89251. .93443
                     /note="assembly_name:Contig255"
     misc_feature    93464. .97169
                     /note="assembly_name:Contig254"
     misc_feature    97190. .102251
                     /note="assembly_name:Contig253"
     misc_feature    102272. .107091
                     /note="assembly_name:Contig252"
     misc_feature    107112. .110480
                     /note="assembly_name:Contig251"
     misc_feature    110501. .113418
                     /note="assembly_name:Contig250"
     misc_feature    113439. .116668
                     /note="assembly_name:Contig249"
     misc_feature    116689. .118403
                     /note="assembly_name:Contig248"
```

```
     misc_feature    118424. .120544
                     /note="assembly_name:Contig247"
     misc_feature    120565. .121993
                     /note="assembly_name:Contig246
                     clone_end:SP6
                     vector_side:left"
     misc_feature    122014. .122930
                     /note="assembly_name:Contig245"
     misc_feature    122951. .124509
                     /note="assembly_name:Contig244"
     misc_feature    124530. .126851
                     /note="assembly_name:Contig243"
     misc_feature    126872. .128496
                     /note="assembly_name:Contig242"
     misc_feature    128517. .129949

  Query Match             82.0%;  Score 192.6;  DB 2;  Length 138860;
  Best Local Similarity   92.3%;  Pred. No. 5e-50;
  Matches  217;  Conservative    0;  Mismatches    9;  Indels    9;  Gaps    1;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108054 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 107995

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||| ||||||||||||||||         |||||||||||||| ||||||||||||||||
Db     107994 GCAACAGTACCAATGGCAGC---------GAGTCTTCCAAGAACCGCACAGTCTCTGGTG 107944

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||||  | ||||||||||||||||||||||||||| || |
Db     107943 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTCC 107884

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| ||||||||||||||||||||||||||||||||||||| |||||
Db     107883 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 107829


Search completed: February 18, 2004, 15:26:04
Job time : 1724.07 secs
```

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 08:17:00 ; Search time 241.401 Seconds
(without alignments)
2627.862 Million cell updates/sec

Title: US-10-026-341A-1
Perfect score: 235
Sequence: 1 atggctggcagatcatctct.....ccacagttccagaccgttga 235

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	235	100.0	235	24	AAD44320	Sp1 gene fragment.
	2	235	100.0	2862	25	AAD51694	Human nucleic acid
	3	235	100.0	3090	22	AAI60335	Human polynucleoti
	4	235	100.0	3289	22	AAI58549	Human polynucleoti
	5	36.4	15.5	2691	23	ABL24299	Drosophila melanog
	6	36.4	15.5	4954	21	AAA63923	DNA encoding a Dro
	7	36.4	15.5	8104	23	ABL24298	Drosophila melanog
	8	33.6	14.3	4745	24	ABS54601	Synechococcus sp.
c	9	33.4	14.2	1024	24	ABN74235	Bovine embryonic g
c	10	33	14.0	1024	25	ABZ83230	Toxicologically re
c	11	33	14.0	2864	22	AAH76458	cDNA corresponding
c	12	33	14.0	2865	24	ABL61798	Colon adenocarcino
c	13	32	13.6	10587	22	AAK82382	Human immune/haema
	14	31.6	13.4	147309	24	ABK49450	Human transporter
c	15	31.4	13.4	289	21	AAA31706	Plant microsatelli
	16	31	13.2	10609	22	AAK70162	Human immune/haema
	17	31	13.2	10631	22	AAK70159	Human immune/haema
	18	31	13.2	48203	22	AAK70161	Human immune/haema
c	19	31	13.2	48203	22	AAK81663	Human immune/haema
	20	31	13.2	48203	22	AAK82628	Human immune/haema
	21	31	13.2	48204	22	AAK70164	Human immune/haema
c	22	31	13.2	48204	22	AAK81666	Human immune/haema
	23	31	13.2	48204	22	AAK82630	Human immune/haema
	24	30.8	13.1	770	23	ABK50302	Potato starch bran
	25	30.8	13.1	3231	17	AAT42632	Class A starch bra
c	26	30.8	13.1	15004	22	AAH27885	Nucleotide sequenc
	27	30.6	13.0	6806	24	ABK97370	Cystatin A (CSTA)
c	28	30.2	12.9	359	22	AAI29247	Colon tumour relat
c	29	30.2	12.9	359	24	ABK45662	cDNA encoding colo
c	30	30.2	12.9	359	25	ABZ33433	Human colon tumour
c	31	30.2	12.9	600	22	AAD06869	Reverse translatio
c	32	30.2	12.9	939	22	AAD06868	Reverse translatio
	33	30.2	12.9	1076	24	ABS76533	cDNA encoding huma
	34	30.2	12.9	1288	19	AAV48113	Nucleotide sequenc
	35	30.2	12.9	1289	19	AAV61202	Full length cDNA s
	36	30.2	12.9	1289	19	AAV58587	Prostate tumour sp
	37	30.2	12.9	1289	22	AAS63558	Human prostate cDN
	38	30.2	12.9	1289	22	AAS10109	Human prostate tum
	39	30.2	12.9	1289	22	AAH93466	Human prostate-spe
	40	30.2	12.9	1289	22	AAH84780	Human prostate-spe
	41	30.2	12.9	1289	22	AAH02531	Prostate tumour an
	42	30.2	12.9	1289	22	AAF86954	Human P503S invent
	43	30.2	12.9	1289	24	ABS71255	Human prostate tum
	44	30.2	12.9	1289	24	ABS58639	Prostate tumour cD
	45	30.2	12.9	1289	24	ABL94930	Human N1-1862 cDNA

ALIGNMENTS

RESULT 1
AAD44320
ID AAD44320 standard; DNA; 235 BP.
XX
AC AAD44320;
XX
DT 13-DEC-2002 (first entry)
XX
DE Sp1 gene fragment.
XX
KW Fibrotic condition; gene expression; cirrhosis; hypertrophic scar;
KW gene therapy; fibrosis; skin disorder; sclerodermic lesion; keloid;
KW trauma; surgery; Sp1; ds.
XX
OS Unidentified.
XX
PN WO200266071-A2.
XX
PD 29-AUG-2002.
XX
PF 21-DEC-2001; 2001WO-US49141.
XX
PR 03-JAN-2001; 2001US-259585P.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI Mauviel A;
XX
DR WPI; 2002-667041/71.
XX
PT Treating a fibrotic condition, e.g. cirrhosis, comprises administering

```
PT   antisense Sp1 or decoy Sp1 oligonucleotides that inhibit transcription
PT   or gene expression of an extra-cellular matrix gene or transforming
PT   growth factor-beta -
XX
PS   Claim 2; Page 38; 38pp; English.
XX
CC   The invention relates to a method of treating a fibrotic condition (the
CC   transcription or gene expression of an extra-cellular matrix (ECM) gene
CC   or transforming growth factor-beta (TGF-beta) is inhibited in a mammal).
CC   The method involves administering an antisense Sp1 or a decoy Sp1
CC   oligonucleotide. The method, antisense Sp1 and decoy Sp1 oligonucleotide
CC   are useful for treating fibrotic conditions e.g. cirrhosis, radiation
CC   induced fibrosis, skin disorders (sclerodermic lesions, hypertrophic
CC   scars, keloids), kidney fibrosis, lung fibrosis and myelofibrosis. They
CC   are also useful for treating or preventing fibrotic scarring following
CC   trauma or surgery. The invention is useful in gene therapy. The present
CC   sequence is Sp1 gene fragment.
XX
SQ   Sequence 235 BP; 60 A; 64 C; 57 G; 54 T; 0 other;

  Query Match             100.0%;  Score 235;  DB 24;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-68;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235


RESULT 2
AAD51694
ID   AAD51694 standard; cDNA; 2862 BP.
XX
AC   AAD51694;
XX
DT   16-APR-2003  (first entry)
XX
DE   Human nucleic acid associated protein (NAAP)-20 encoding cDNA.
XX
KW   Human; nucleic acid associated protein; NAAP; cancer; atherosclerosis;
KW   cell proliferative disorder; neurological disorder; Huntington's disease;
KW   epilepsy; stroke; immune disorder; acquired immune deficiency syndrome;
KW   AIDS; inflammatory disorder; allergy; developmental disorder; infection;
KW   hypothyroidism; cushing's syndrome; gene therapy; cytostatic; nootropic;
KW   anticonvulsant; neuroprotective; cerebroprotective; thyromimetic; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             107..2464
FT                   /*tag=  a
FT                   /product= "Human NAAP-20 protein"
FT   sig_peptide     107..325
FT                   /*tag=  b
FT   mat_peptide     326..2461
FT                   /*tag=  c
FT                   /product= "Human mature NAAP-20 protein"
XX
PN   WO200299115-A2.
XX
PD   12-DEC-2002.
XX
PF   31-MAY-2002; 2002WO-US17050.
XX
PR   01-JUN-2001; 2001US-295359P.
PR   08-JUN-2001; 2001US-296878P.
PR   08-JUN-2001; 2001US-297222P.
PR   15-JUN-2001; 2001US-298615P.
PR   15-JUN-2001; 2001US-298665P.
PR   15-JUN-2001; 2001US-298693P.
PR   21-JUN-2001; 2001US-300176P.
PR   19-APR-2002; 2002US-373891P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
PA   (YUEH/) YUE H.
XX
PI   Yue H,  Tang YT,  Baughn MR,  Becha SD,  Warren BA,  Walia NK,  Lal PG;
PI   Lee EA,  Hafalia AJA,  Richardson TW,  Griffin JA,  Emerling BM;
PI   Ramkumar J,  Yue H,  Swarnakar A,  Tran B,  Li JX,  Yao MG,  Yang J;
PI   Ison CH,  Forsythe IJ,  Honchell CD,  Arvizu CS,  Elliott VS,  Lu Y;
PI   Ding L,  Luo W,  Wang YE,  Burford N,  Borowsky ML,  Nguyen DB;
PI   Chinn AM,  Kable AE;
XX
DR   WPI; 2003-140626/13.
DR   P-PSDB; AAE33780.
XX
PT   New human nucleic acid associated proteins (NAAP), useful for
PT   diagnosing, treating and preventing diseases or conditions associated
PT   with the aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis,
PT   epilepsy, or infections  -
XX
PS   Claim 105; Column 247-248; 257pp; English.
XX
CC   The present invention relates to human nucleic acid associated proteins
CC   (NAAP) and polynucleotides encoding such proteins. NAAP sequences are
CC   useful in diagnosing, treating and preventing diseases or conditions
CC   associated with the decreased expression or overexpression of NAAP such
CC   as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC   epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC   allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC   disorders or infections. They are also used in gene therapy. The present
CC   sequence is human NAAP-20 cDNA.
XX
SQ   Sequence 2862 BP; 753 A; 807 C; 712 G; 590 T; 0 other;

  Query Match             100.0%;  Score 235;  DB 25;  Length 2862;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-68;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 479

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 539

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 599

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 654


RESULT 3
AAI60335
ID   AAI60335 standard; cDNA; 3090 BP.
XX
AC   AAI60335;
XX
DT   22-OCT-2001  (first entry)
```

```
XX
DE   Human polynucleotide SEQ ID NO 4324.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM41179.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 4324; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynuclectide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 3090 BP; 814 A; 852 C; 752 G; 672 T; 0 other;

  Query Match           100.0%;  Score 235;  DB 22;  Length 3090;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-68;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        411 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 470

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        471 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 530

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        531 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 590

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        591 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 645


RESULT 4
AAI58549
ID   AAI58549 standard; cDNA; 3289 BP.
XX
AC   AAI58549;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 752.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM39393.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 752; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
```

```
SQ   Sequence 3289 BP; 867 A; 894 C; 786 G; 742 T; 0 other;

  Query Match           100.0%;  Score 235;  DB 22;  Length 3289;
  Best Local Similarity  100.0%;  Pred. No. 4.8e-68;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 573

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 633

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 693

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 748


RESULT 5
ABL24299
ID   ABL24299 standard; DNA; 2691 BP.
XX
AC   ABL24299;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 24370.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 24370; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2691 BP; 689 A; 883 C; 748 G; 371 T; 0 other;

  Query Match           15.5%;  Score 36.4;  DB 23;  Length 2691;
  Best Local Similarity  54.5%;  Pred. No. 0.12;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         30 GGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGCAATGGCAGT 89
              || | | || || | | | | | | |||||| | |||||||||   | ||||
Db        235 GGCGGCAGCCATAGCAAGCATCTGCGCATCAGCAGCACTTCCAATGGCAAGCACGGCAAA 294

Qy         90 GAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTGCCGCTCCCAAC 149
              | ||   | | | |  | | | |  | || |||| |||| ||| ||| |||
Db        295 TACTCAAATATGCAGCAGCAACTGCCGCAGGATGAGGATGTGGTGGATGCGGCTGCCACG 354

Qy        150 TTACAGAACCAGCA 163
              | ||| | |||||
Db        355 ATGCAGCAGCAGCA 368


RESULT 6
AAA63923
ID   AAA63923 standard; DNA; 4954 BP.
XX
AC   AAA63923;
XX
DT   04-DEC-2000  (first entry)
XX
DE   DNA encoding a Drosophila naked cuticle polypeptide.
XX
KW   Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signalling;
KW   gene therapy; Nkd defect; cancer; ss.
XX
OS   Drosophila melanogaster.
XX
FH   Key             Location/Qualifiers
FT   CDS             966..3752
FT                   /*tag=  a
FT                   /product= "naked cuticle polypeptide"
XX
PN   WO200049034-A1.
XX
PD   24-AUG-2000.
XX
PF   17-FEB-2000; 2000WO-US04188.
XX
PR   17-FEB-1999;   99US-0120646.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Scott M,  Zeng W,  Wharton K;
XX
DR   WPI; 2000-571967/53.
DR   P-PSDB; AAB08214.
XX
PT   An isolated nucleic acid molecule useful for analyzing (genetic
PT   predisposition to) a disease state and for therapeutic purposes e.g.
PT   treatment of cancer comprises a sequence encoding a naked cuticle
PT   protein -
XX
PS   Claim 3; Page 36-40; 58pp; English.
XX
CC   The present sequence encodes a Nkd (naked cuticle) polypeptide. In
CC   Drosophila, Nkd is a segment-polarity gene whose expression is
CC   induced by Wnt signalling. The Nkd polypeptide acts to antagonize Wnt
CC   signalling. Nkd may link ion fluxes to the regulation of Wnt signal
CC   potency, duration or distribution. The Nkd polynucleotides can be
CC   used for identifying homologous or related proteins, to modulate the
CC   expression or function of Nkd polypeptides, and in studying associated
CC   physiological pathways. Nkd polynucleotides can also be used in gene
CC   therapy to treat disorders associated with Nkd defects.  They may
CC   also be used for therapeutic purposes e.g. treatment of cancer.
XX
SQ   Sequence 4954 BP; 1534 A; 1330 C; 1154 G; 936 T; 0 other;

  Query Match           15.5%;  Score 36.4;  DB 21;  Length 4954;
```

```
  Best Local Similarity   54.5%;  Pred. No. 0.16;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          30 GGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGCAATGGCAGT 89
               || |  | || || |    |    |  |  |||||| | |||||||||    | ||||
Db        1296 GGCGGCAGCCATAGCAAGCATCTGCGCATCAGCAGCACTTCCAATGGCAAGCACGGCAAA 1355

Qy          90 GAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTGCCGCTCCCAAC 149
               | ||     | | | |    | |  |    || |||| |||| ||| ||| |||
Db        1356 TACTCAAATATGCAGCAGCAACTGCCGCAGGATGAGGATGTGGTGGATGCGGCTGCCACG 1415

Qy         150 TTACAGAACCAGCA 163
               | ||| | |||||
Db        1416 ATGCAGCAGCAGCA 1429


RESULT 7
ABL24298
ID   ABL24298 standard; DNA; 8104 BP.
XX
AC   ABL24298;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 24367.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 24367; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 8104 BP; 2422 A; 1863 C; 1717 G; 2102 T; 0 other;

  Query Match             15.5%;  Score 36.4;  DB 23;  Length 8104;
  Best Local Similarity   54.5%;  Pred. No. 0.2;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          30 GGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGCAATGGCAGT 89
               || |  | || || |    |    |  |  |||||| | |||||||||    | ||||
Db        1235 GGCGGCAGCCATAGCAAGCATCTGCGCATCAGCAGCACTTCCAATGGCAAGCACGGCAAA 1294

Qy          90 GAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTGCCGCTCCCAAC 149
               | ||     | | | |    | |  |    || |||| |||| ||| ||| |||
Db        1295 TACTCAAATATGCAGCAGCAACTGCCGCAGGATGAGGATGTGGTGGATGCGGCTGCCACG 1354

Qy         150 TTACAGAACCAGCA 163
               | ||| | |||||
Db        1355 ATGCAGCAGCAGCA 1368


RESULT 8
ABS54601
ID   ABS54601 standard; DNA; 4745 BP.
XX
AC   ABS54601;
XX
DT   28-NOV-2002  (first entry)
XX
DE   Synechococcus sp. isoamylase gene.
XX
KW   ds; gene; isoamylase; PCR; gene cloning.
XX
OS   Synechococcus sp. strain pcc7942.
XX
FH   Key             Location/Qualifiers
FT   CDS             1647..3731
FT                   /*tag=  a
FT                   /product= "Isoamylase"
XX
PN   JP2002262877-A.
XX
PD   17-SEP-2002.
XX
PF   07-MAR-2001; 2001JP-0064133.
XX
PR   07-MAR-2001; 2001JP-0064133.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA   (UYTY ) UNIV TOKYO.
XX
DR   WPI; 2002-703285/76.
DR   P-PSDB; ABG70853.
XX
PT   A new isoamylase, cloning a gene, a primer, an isoamylase, a gene, an
PT   organism  -
XX
PS   Claim 7; Page 12-15; 22pp; Japanese.
XX
CC   The invention relates to a method for cloning a gene encoding isoamylase
CC   based on a DNA amplified by a PCR from a DNA of a genome or a library
CC   using the PCR primers appearing as ABS54602 and ABS54603. The primers may
CC   also amplify a base sequence in which part of the bases is replaced,
CC   deleted and/or added in it by a PCR. Also included are an isoamylase
CC   appearing as ABG70853 or an amino acid sequence in which part of the
CC   amino acids is deleted, replaced or added in it, a gene encoding the
CC   above isoamylase, and an organism in which the above gene is introduced
CC   to an organism in which isoamylase is deleted. The method is used for
CC   cloning a gene encoding isoamylase. The present sequence is the
CC   isoamylase gene from Synechococcus isolated using the primers of
CC   the invention.
XX
SQ   Sequence 4745 BP; 930 A; 1385 C; 1258 G; 1171 T; 1 other;

  Query Match             14.3%;  Score 33.6;  DB 24;  Length 4745;
  Best Local Similarity   49.2%;  Pred. No. 1.4;
  Matches   87;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy          56 GAGTGGCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTC 115
               |||| || ||||         ||||| |||| |||| |  | | |  |  ||    | |
Db        4251 GAGTAGCGGCAGGCGTGGCGGCAGGAATGCAAGTGTTTGNTACGCGGGTTCCATGCCCGC 4310

Qy         116 TGGTGGGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGG 175
               | |  | | | | |  | | |||| || ||  ||  |  |  || | | ||| |  |
```

```
Db       4311 TTGGCGTCTGCAAGAAGCCGGTGCCCATCTCATTTTTGACGATATGCGACTGCTGCCCAG 4370

Qy        176 ACTACCTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGT 232
              |||||     |  |  ||| ||   | | |    || | | ||  |   |  |  ||| |
Db       4371 TCTGCTCCAATCGTCGCCAAAAGATAACTCCACAGCATTGCCCAATCCCTAACCCCT 4427


RESULT 9
ABN74235/c
ID   ABN74235 standard; cDNA; 1024 BP.
XX
AC   ABN74235;
XX
DT   03-JUL-2002  (first entry)
XX
DE   Bovine embryonic germ (EG) cell cDNA EST #286.
XX
KW   Bovine; Bos taurus; EST; expressed sequence tag; totipotence;
KW   development; gene; ss.
XX
OS   Bos taurus.
XX
PN   WO200194550-A2.
XX
PD   13-DEC-2001.
XX
PF   07-JUN-2001; 2001WO-US18576.
XX
PR   07-JUN-2000; 2000US-209874P.
PR   06-JUN-2001; 2001US-0876143.
XX
PA   (INFI-) INFIGEN INC.
XX
PI   Eilertsen KJ,  Pfister-Genskow M,  Childs L;
XX
DR   WPI; 2002-351289/38.
XX
PT   An expressed sequence tag (EST), the expression of which, or its
PT   complementary sequence, in a cell identifies the cell as a
PT   developmentally competent or incompetent cell -
XX
PS   Example 16; Page 364; 584pp; English.
XX
CC   The present invention describes an expressed sequence tag (EST), where
CC   the EST is an isolated, enriched, or purified nucleic acid sequence
CC   representing all or part of a gene, the expression of which, or its
CC   complementary sequence, in a cell identifies the cell as a
CC   developmentally competent or incompetent cell. Molecules which induce
CC   developmental competence in a cell line are useful for inducing
CC   totipotence in one or more cells. Molecules which induce developmental
CC   incompetence in a cell line are useful for preventing a full term
CC   pregnancy in an animal and inhibiting totipotence. The molecules are
CC   also useful for treating a disease in an animal by inducing development
CC   of one or more cells of the animal into a specific cell type. The
CC   present sequence represents a bovine EST which is given in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1024 BP; 187 A; 274 C; 282 G; 227 T; 54 other;

  Query Match           14.2%;  Score 33.4;  DB 24;  Length 1024;
  Best Local Similarity   54.5%;  Pred. No. 0.84;
  Matches   67;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy         19 CTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCA 78
              || ||  ||  ||| | |||     |||   ||| ||| ||||||||   |   | |
Db        570 CTCCCCGCTGCGGGTCCACCTGGCGCTCGGGGGACCAGGGTGGCAGCTTGAGGGACGAGC 511

Qy         79 GCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTG 138
              ||||  | ||  || | |    ||||  | ||| ||| |    |  | | || |||
Db        510 GCAAGAGGAGGAAGCCCTGGTGGAATGACCCAGGCTCAGAAACGTGGCACGTGGTGAGCC 451

Qy        139 CCG 141
              |||
Db        450 CCG 448


RESULT 10
ABZ83230/c
ID   ABZ83230 standard; cDNA; 1024 BP.
XX
AC   ABZ83230;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Toxicologically relevant human nucleotide sequence #389.
XX
KW   Toxicologically relevant gene; toxicological response; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003016500-A2.
XX
PD   27-FEB-2003.
XX
PF   16-AUG-2002; 2002WO-US26514.
XX
PR   16-AUG-2001; 2001US-313080P.
XX
PA   (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
XX
PI   Neft RE,  Dunn RT,  Adkins K,  Pickett GG,  Kier LD,  Schmeiser K;
PI   Alen P;
XX
DR   WPI; 2003-268322/26.
XX
PT   Determining a toxicological response to an agent, useful for screening
PT   of drugs, comprises comparing the expression profile of one or more
PT   human toxic response genes to a reference gene expression profile
PT   indicative of toxicity -
XX
PS   Claim 1; Page 140; 455pp; English.
XX
CC   The present invention describes a method (M1) for determining a
CC   toxicological response to an agent, which comprises comparing the
CC   expression profile of one or more human toxic response genes to a
CC   reference gene expression profile indicative of toxicity, and so
CC   determining the presence of a toxic response to the agent. Also
CC   described: (1) an array comprising one or more polynucleotides selected
CC   from the genes corresponding to the partial sequences given in ABZ82842
CC   to ABZ84764, or their fragments of at least 20 nucleotides, or
CC   homologues; and (2) determining if a gene putatively identified to be a
CC   toxic response gene plays a role on toxic response pathways by
CC   determining the expression profile of the gene after exposure of cells
CC   or a human subject to a known toxic pharmaceutical or industrial agent,
CC   comprising: (a) exposing cells to an agent or isolating cells from a
CC   human subject who was exposed to an agent; (b) obtaining the test gene
CC   expression profile for a putatively identified toxic response gene after
CC   exposure to a known toxic pharmaceutical or industrial agent; and
CC   (c) comparing the test profile to the expression profile of a gene with
CC   a similar function or comparing the test profile to the expression
CC   profile of that gene after exposure to other known toxic compounds. The
CC   methods are useful for predicting and determining toxicological responses
CC   on a cellular, organ or system level. The arrays comprising the human
CC   genes are useful for toxicological screening of drugs, pharmaceutical
CC   compounds and chemicals.
XX
SQ   Sequence 1024 BP; 271 A; 251 C; 289 G; 213 T; 0 other;

  Query Match           14.0%;  Score 33;  DB 25;  Length 1024;
  Best Local Similarity   57.1%;  Pred. No. 1.1;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         21 TCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGC 80
              ||||| || ||| |  |  |  || | | ||      ||  ||| ||| ||||  |||||
Db        134 TCCTCGTCCGGGCCCTCGGCACCCGCCACGGCCAGCAGCTGCACCAGCCCCAAGAGCAGC 75
```

```
Qy          81 AATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAG 125
               |   ||||  || ||| |  |   | ||  ||  | || | | ||
Db          74 AGGAGCAGGAAGGCTTTCCGGGGCCTCATGGTAGCGGGGGCGGAG 30
```

RESULT 11
AAH76458/c
ID AAH76458 standard; cDNA; 2864 BP.
XX
AC AAH76458;
XX
DT 22-OCT-2001 (first entry)
XX
DE cDNA corresponding to human IFN-alpha induced gene encoding ERP-70.
XX
KW Human; interferon-alpha induced gene; type I interferon treatment;
KW chronic viral hepatitis; relapsing remitting multiple sclerosis;
KW neoplastic disease; IFN-alpha; interferon-alpha; ERP-70; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   46..1983
FT   /*tag= a
FT   /product= "ERP-70"
XX
PN WO200159155-A2.
XX
PD 16-AUG-2001.
XX
PF 09-FEB-2001; 2001WO-GB00578.
XX
PR 11-FEB-2000; 2000GB-0003203.
PR 11-FEB-2000; 2000GB-0003204.
PR 11-FEB-2000; 2000GB-0003205.
PR 11-FEB-2000; 2000GB-0003206.
PR 11-FEB-2000; 2000GB-0003207.
PR 11-FEB-2000; 2000GB-0003208.
PR 11-FEB-2000; 2000GB-0003210.
PR 11-FEB-2000; 2000GB-0003212.
PR 11-FEB-2000; 2000GB-0003213.
PR 11-FEB-2000; 2000GB-0003215.
PR 11-FEB-2000; 2000GB-0003220.
PR 11-FEB-2000; 2000GB-0003221.
PR 11-FEB-2000; 2000GB-0003222.
PR 17-FEB-2000; 2000GB-0003768.
XX
PA (PHAR-) PHARMA PACIFIC PTY LTD.
XX
PI Meritet J, Dron M, Tovey MG;
XX
DR WPI; 2001-483570/52.
DR P-PSDB; AAG66531.
XX
PT Predicting responsiveness of a patient to treatment with a type I
PT interferon comprising determining the level of induced proteins after
PT treatment with a type I interferon, -
XX
PS Example 1; Page 54-57; 133pp; English.
XX
CC The invention relates to a method for predicting responsiveness of a
CC patient to treatment with a type I interferon. The method comprises
CC determining the level of one or more proteins with a 646, 164, 126, 598,
CC 98, 177, 761, 361, 941, 657, 817, 429, 473, 399, 285 or 303 amino acid
CC sequence fully defined in the specification after treatment with a
CC type I interferon. The method allows a physician to determine whether
CC a patient suffering from chronic viral hepatitis, neoplastic disease
CC or relapsing remitting multiple sclerosis will respond favourably to
CC Type I interferon treatment via oromucosal administration. The
CC present sequence is a cDNA corresponding to an interferon-alpha
CC induced gene that encodes one of the polypeptides listed above.
XX
SQ Sequence 2864 BP; 786 A; 643 C; 729 G; 706 T; 0 other;

```
  Query Match             14.0%;  Score 33;  DB 22;  Length 2864;
  Best Local Similarity   57.1%;  Pred. No. 1.8;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy          21 TCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGC 80
               ||||| || ||| |  |  |  || | | ||      ||  ||| |||  |||| |||||
Db         134 TCCTCGTCCGGGCCCTCGGCACCCGCCACGGCCAGCAGCTGCACCAGCCCCAAGAGCAGC 75

Qy          81 AATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAG 125
               |   ||||  || ||| |  |   | ||  ||  | || | | ||
Db          74 AGGAGCAGGAAGGCTTTCCGGGGCCTCATGGTAGCGGGGGCGGAG 30
```

RESULT 12
ABL61798/c
ID ABL61798 standard; DNA; 2865 BP.
XX
AC ABL61798;
XX
DT 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:135.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.

PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 135; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 2865 BP; 786 A; 643 C; 729 G; 707 T; 0 other;

Query Match 14.0%; Score 33; DB 24; Length 2865;
Best Local Similarity 57.1%; Pred. No. 1.8;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy        21 TCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGC 80
             ||||| || ||| |  |  |  || | | ||      ||  ||| |||  ||||  |||||
Db       134 TCCTCGTCCGGGCCCTCGGCACCCGCCACGGCCAGCAGCTGCACCAGCCCCAAGAGCAGC 75

Qy        81 AATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAG 125
             |   ||||  || ||| |  |   | ||  ||  | || | | ||
Db        74 AGGAGCAGGAAGGCTTTCCGGGGCCTCATGGTAGCGGGGGCGGAG 30
```

RESULT 13
AAK82382/c
ID AAK82382 standard; DNA; 10587 BP.
XX
AC AAK82382;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37194.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.

XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.

```
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 37194; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 10587 BP; 2923 A; 2113 C; 2189 G; 3362 T; 0 other;

  Query Match             13.6%;  Score 32;  DB 22;  Length 10587;
  Best Local Similarity   52.2%;  Pred. No. 6.6;
  Matches   71;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy         70 CCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATG 129
Db       7286 CAAATGAGAGCCCTACAGGTCAATCTGGCATTGCTGGCACTGTCGTCCATGGGTAGCACT 7227

Qy        130 TTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTACCTGGAGTGA 189
Db       7226 GCTTAGTCCCCTGTCCCCACCCACAAAACCACTTAGTTTTACTTGAATCACATGAATATA 7167

Qy        190 TGCCTAATATTCAGTA 205
Db       7166 CTCCATGTATTTTGAA 7151


RESULT 14
ABK49450
ID   ABK49450 standard; DNA; 147309 BP.
XX
AC   ABK49450;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human transporter genomic DNA sequence.
XX
KW   Human; transporter protein; therapeutic target; query sequence;
KW   database search; single nucleotide polymorphism; SNP; chromosome 5;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            2002..2778
FT                   /*tag=  a
FT                   /number= 1
FT   intron          2779..43601
FT                   /*tag=  b
FT                   /number= 1
FT   exon            43602..43784
FT                   /*tag=  c
FT                   /number= 2
FT   intron          43785..90350
FT                   /*tag=  d
FT                   /number= 2
FT   exon            90351..90354
FT                   /*tag=  e
FT                   /number= 3
```

```
FT   intron          90355..92190
FT                   /*tag= f
FT                   /number= 3
FT   exon            92191..92403
FT                   /*tag= g
FT                   /number= 4
FT   intron          92404..106742
FT                   /*tag= h
FT                   /number= 4
FT   exon            106743..106976
FT                   /*tag= i
FT                   /number= 5
FT   intron          106977..110890
FT                   /*tag= j
FT                   /number= 5
FT   exon            110891..111155
FT                   /*tag= k
FT                   /number= 6
FT   intron          111156..127189
FT                   /*tag= l
FT                   /number= 6
FT   exon            127190..127385
FT                   /*tag= m
FT                   /number= 7
FT   intron          127386..132484
FT                   /*tag= n
FT                   /number= 7
FT   exon            132485..132706
FT                   /*tag= o
FT                   /number= 8
FT   intron          132707..133385
FT                   /*tag= p
FT                   /number= 8
FT   exon            133386..133466
FT                   /*tag= q
FT                   /number= 9
FT   intron          133467..142467
FT                   /*tag= r
FT                   /number= 9
FT   exon            142468..142543
FT                   /*tag= s
FT                   /number= 10
FT   intron          142544..145139
FT                   /*tag= t
FT                   /number= 10
FT   exon            145140..145306
FT                   /*tag= u
FT                   /number= 11
FT   variation       replace (3356,T)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (3396,C)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (3493,A)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (3778..3780, GC)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (4063,G)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (4628,G)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (4702,A)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (5235,C)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (5638,A)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (6248,C)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (7845,A)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (8690,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10023,A)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10049,T)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10300,A)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (10842,C)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (12545,G)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (13192,T)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14227,C)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14304,G)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14535,C)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14580,A)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14594,G)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (14836,C)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (15578,A)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (16079,G)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (16118,G)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (16643,A)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (16712,C)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (17201,C)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (17228,T)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (17893..17895, AG)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (17895..17897, GG)
FT                   /*tag= bb
```

```
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (19102,A)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (21708,G)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (22641,A)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (24758,G)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (25059,C)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (26084..26086, CAAC)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (26318,G)
FT                   /*tag=  bi
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (27203..27205, CTTG)
FT                   /*tag=  bj

  Query Match             13.4%;  Score 31.6;  DB 24;  Length 147309;
  Best Local Similarity   58.5%;  Pred. No. 27;
  Matches   55;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy        130 TTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTACCTGGAGTGA 189
              |||  |||||| |||||| |      ||  |||   ||||| |  | |  | || ||||||
Db      97837 TTGCTGCTGCCCCTCCCACCCCCATGACTCAGTCTGTTCTCAATGCAGGAGCTAGAGTGA 97896

Qy        190 TGCCTAATATTCAGTATCAAGTAATCCCACAGTT 223
              | | |   | |   ||  |||  || ||  |||
Db      97897 TCCATTTAAAACGTAATTGAGTCTTCTCATTGTT 97930


RESULT 15
AAA31706/c
ID   AAA31706 standard; DNA; 289 BP.
XX
AC   AAA31706;
XX
DT   05-JUL-2000  (first entry)
XX
DE   Plant microsatellite marker #667.
XX
KW   Plant microsatellite sequence; core repeat sequence; detection; probe;
KW   DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW   variety identification; genetic variability evaluation; primer; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO9967421-A1.
XX
PD   29-DEC-1999.
XX
PF   25-JUN-1999;   99WO-NZ00092.
XX
PR   25-JUN-1998;   98US-0105307.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Havukkala IJ,  Bloksberg LN,  Glenn M;
XX
DR   WPI; 2000-116958/10.
XX
PT   New plant microsatellite markers and associated flanking species for
PT   the detection of polymorphic genetic markers -
XX
PS   Claim 1; Page 273; 392pp; English.
XX
CC   Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC   and associated flanking species. The sequences comprise a central core
CC   repeat sequence, especially selected from the sequences AAA32094-A32096
CC   with left and right flanking sequences. The polynucleotide sequences
CC   can be used in the detection of DNA polymorphisms, in genome mapping,
CC   in physical mapping, in positional cloning of genes, in variety
CC   identification and in evaluation of genetic variability within and
CC   between plant tissues, populations, cultivars, species and species
CC   groups. They may also be used to design hybridization probes for
CC   oligonucleotide fingerprinting and library screening and to design
CC   primers for microsatellite-primed PCR. Microsatellite markers are
CC   useful to locate specific economically useful genes in plant genomes.
XX
SQ   Sequence 289 BP; 59 A; 67 C; 82 G; 81 T; 0 other;

  Query Match             13.4%;  Score 31.4;  DB 21;  Length 289;
  Best Local Similarity   64.4%;  Pred. No. 2.3;
  Matches   47;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         16 TCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATG 75
              || || |   ||||| ||| | | ||  ||| || |   | ||  ||||||| | ||
Db        105 TCCCTACAATCTCTGTGGCCATCTCTCTCTCTAACGGTGACAGCAGCAGCAGCAGCAGCA 46

Qy         76 GCAGCAATGGCAG 88
              ||||||  |||||
Db         45 GCAGCAGCGGCAG 33


Search completed: February 18, 2004, 14:54:30
Job time : 243.401 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 14:29:50 ; Search time 62.179 Seconds
(without alignments)
1668.169 Million cell updates/sec

Title: US-10-026-341A-1
Perfect score: 235
Sequence: 1 atggctggcagatcatctct.....ccacagttccagaccgttga 235

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	235	100.0	3289	4	US-09-620-312D-436	Sequence 436, App
	2	31.2	13.3	2073	4	US-09-221-017B-894	Sequence 894, App
	3	30.2	12.9	1289	3	US-09-020-956-111	Sequence 111, App
	4	30.2	12.9	1289	3	US-09-030-607-111	Sequence 111, App
	5	30.2	12.9	1289	4	US-09-439-313-111	Sequence 111, App
	6	30.2	12.9	1289	4	US-09-352-616A-111	Sequence 111, App
	7	30.2	12.9	1289	4	US-09-232-149A-111	Sequence 111, App
c	8	29.6	12.6	621	4	US-09-328-352-2777	Sequence 2777, Ap
	9	29.2	12.4	7286	4	US-08-793-273C-1	Sequence 1, Appli
	10	29.2	12.4	7286	5	PCT-US95-11684-1	Sequence 1, Appli
c	11	29	12.3	354	2	US-08-958-201-1	Sequence 1, Appli
c	12	28.8	12.3	1030	4	US-09-831-206-1	Sequence 1, Appli
	13	28.8	12.3	9751	3	US-09-238-303-7	Sequence 7, Appli
	14	28.8	12.3	9751	4	US-09-946-239-7	Sequence 7, Appli
	15	28.4	12.1	3269	4	US-09-634-238-409	Sequence 409, App
c	16	28	11.9	354	2	US-08-958-201-3	Sequence 3, Appli
c	17	27.8	11.8	1000	3	US-09-018-584A-38	Sequence 38, Appl
	18	27.6	11.7	1286	1	US-08-218-943-5	Sequence 5, Appli
	19	27.6	11.7	3023	4	US-09-203-453-4	Sequence 4, Appli
	20	27.6	11.7	3023	4	US-09-900-236-4	Sequence 4, Appli
	21	27.6	11.7	3074	3	US-09-087-277-1	Sequence 1, Appli
	22	27.6	11.7	3074	4	US-09-658-499-1	Sequence 1, Appli
c	23	27.6	11.7	3530	3	US-08-704-711A-10	Sequence 10, Appl
c	24	27.6	11.7	3530	4	US-09-521-220-10	Sequence 10, Appl
c	25	27.4	11.7	847	1	US-07-934-393B-5	Sequence 5, Appli
	26	27.4	11.7	985	3	US-09-160-246-9	Sequence 9, Appli
	27	27.4	11.7	985	3	US-09-160-246-13	Sequence 13, Appl
c	28	27.4	11.7	1155	4	US-09-163-951-15	Sequence 15, Appl
c	29	27.4	11.7	1240	4	US-09-163-951-14	Sequence 14, Appl
	30	27.2	11.6	279	4	US-09-313-294A-4554	Sequence 4554, Ap
c	31	27.2	11.6	996	1	US-08-671-525B-7	Sequence 7, Appli
c	32	27.2	11.6	996	1	US-08-672-109B-7	Sequence 7, Appli
c	33	27.2	11.6	996	1	US-08-842-045-7	Sequence 7, Appli
c	34	27.2	11.6	996	2	US-08-842-238-7	Sequence 7, Appli
c	35	27.2	11.6	996	3	US-08-629-335B-7	Sequence 7, Appli
c	36	27.2	11.6	999	3	US-08-870-511-7	Sequence 7, Appli
c	37	27.2	11.6	999	3	US-08-870-511-9	Sequence 9, Appli
c	38	27.2	11.6	999	3	US-08-870-511-11	Sequence 11, Appl
c	39	27.2	11.6	999	4	US-09-016-434-1277	Sequence 1277, Ap
c	40	27.2	11.6	1671	2	US-08-662-560-1	Sequence 1, Appli
c	41	27.2	11.6	1671	2	US-08-780-749A-5	Sequence 5, Appli
c	42	27.2	11.6	1671	3	US-08-706-281A-15	Sequence 15, Appl
c	43	27.2	11.6	1671	3	US-09-097-231-15	Sequence 15, Appl
c	44	27.2	11.6	1671	3	US-08-870-511-5	Sequence 5, Appli
c	45	27.2	11.6	1671	4	US-09-353-099-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-620-312D-436
; Sequence 436, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 436
; LENGTH: 3289
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(2558)
US-09-620-312D-436

Query Match 100.0%; Score 235; DB 4; Length 3289;
Best Local Similarity 100.0%; Pred. No. 1.5e-73;
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 573
```

```
Qy          61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         574 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 633

Qy         121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         634 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 693

Qy         181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         694 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 748
```

RESULT 2
US-09-221-017B-894
; Sequence 894, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 894:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2073 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...2073
US-09-221-017B-894

Query Match 13.3%; Score 31.2; DB 4; Length 2073;
Best Local Similarity 50.7%; Pred. No. 0.45;
Matches 75; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy           4 GCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCA 63
               || | || | |  || |||||||  ||  | |||  || |    |  ||| |
Db         394 GCCGTCAAACCCGTTCGTCCTCCTTCGGCAATGCCCGATCCGCTCTCCACATAAGTAGTT 453

Qy          64 GCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGC 123
                 |||      || |     ||||||  ||| |||  |  || |||  | |  |  ||| |
Db         454 ACCGTGGAGGTGCCCAGTATGGCACCGAGGCTTGTACCGATGGCATCGGCCATCATGGCC 513

Qy         124 AGTATGTTGTGGCTGCCGCTCCCAACTT 151
                 |  || | |   || |  ||||  ||
Db         514 TGCTTGATATTCTTGACATTCCCTTTTT 541
```

RESULT 3
US-09-020-956-111
; Sequence 111, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillin, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-020-956-111

Query Match 12.9%; Score 30.2; DB 3; Length 1289;
Best Local Similarity 55.1%; Pred. No. 0.8;
Matches 59; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
Qy          17 CTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGG 76
               | ||||  ||||||   ||||  || || ||  ||||| | | |  | |||          |
Db         852 CACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGCACCCTGGCAAG 911

Qy          77 CAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGC 123
               |||||  ||   | | |        || |||  ||| | ||   |||||
Db         912 CAGCAGTGATTGGGGGAGGGGACAGGATCTAACAATGTCACTTGGGC 958
```

```
RESULT 4
US-09-030-607-111
; Sequence 111, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  111:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1289 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-111

  Query Match             12.9%;  Score 30.2;  DB 3;  Length 1289;
  Best Local Similarity   55.1%;  Pred. No. 0.8;
  Matches   59;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy         17 CTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGG 76
              | ||||  ||||||   ||||  || || ||  ||||| | |  |  |||          |
Db        852 CACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGCACCCTGGCAAG 911

Qy         77 CAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGC 123
              |||||  ||   | | |        || |||  ||| | ||   |||||
Db        912 CAGCAGTGATTGGGGGAGGGGACAGGATCTAACAATGTCACTTGGGC 958


RESULT 5
US-09-439-313-111
; Sequence 111, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 111
;   LENGTH: 1289
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-439-313-111

  Query Match             12.9%;  Score 30.2;  DB 4;  Length 1289;
  Best Local Similarity   55.1%;  Pred. No. 0.8;
  Matches   59;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy         17 CTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGG 76
              | ||||  ||||||   ||||  || || ||  ||||| | |  |  |||          |
Db        852 CACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGCACCCTGGCAAG 911

Qy         77 CAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGC 123
              |||||  ||   | | |        || |||  ||| | ||   |||||
Db        912 CAGCAGTGATTGGGGGAGGGGACAGGATCTAACAATGTCACTTGGGC 958


RESULT 6
US-09-352-616A-111
; Sequence 111, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang, Yuqui
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Mitcham, Jennifer Lynn
;  TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF PROSTATE CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE:  210121.427C8
;  CURRENT APPLICATION NUMBER: US/09/352,616A
;  CURRENT FILING DATE:  1999-07-13
;  NUMBER OF SEQ ID NOS:  472
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 111
;   LENGTH: 1289
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-352-616A-111

  Query Match             12.9%;  Score 30.2;  DB 4;  Length 1289;
  Best Local Similarity   55.1%;  Pred. No. 0.8;
  Matches   59;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy         17 CTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGG 76
              | ||||  ||||||   ||||  || || ||  ||||| | |  |  |||          |
Db        852 CACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGCACCCTGGCAAG 911

Qy         77 CAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGC 123
              |||||  ||   | | |        || |||  ||| | ||   |||||
Db        912 CAGCAGTGATTGGGGGAGGGGACAGGATCTAACAATGTCACTTGGGC 958


RESULT 7
US-09-232-149A-111
; Sequence 111, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer Lynn
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
;  TITLE OF INVENTION:  CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.427C6
;  CURRENT APPLICATION NUMBER: US/09/232,149A
```

```
;  CURRENT FILING DATE:  1999-01-15
;  NUMBER OF SEQ ID NOS: 338
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 111
;    LENGTH: 1289
;    TYPE: DNA
;    ORGANISM: Homo sapien
US-09-232-149A-111

  Query Match             12.9%;  Score 30.2;  DB 4;  Length 1289;
  Best Local Similarity   55.1%;  Pred. No. 0.8;
  Matches   59;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy         17 CTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGG 76
              | ||||  ||||||   ||||  || || ||  ||||| | |  |  |||          |
Db        852 CACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGCACCCTGGCAAG 911

Qy         77 CAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGC 123
              |||||  ||   | | |       || |||  ||| | ||  |||||
Db        912 CAGCAGTGATTGGGGGAGGGGACAGGATCTAACAATGTCACTTGGGC 958


RESULT 8
US-09-328-352-2777/c
; Sequence 2777, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2777
;    LENGTH: 621
;    TYPE: DNA
;    ORGANISM: Acinetobacter baumannii
US-09-328-352-2777

  Query Match             12.6%;  Score 29.6;  DB 4;  Length 621;
  Best Local Similarity   68.3%;  Pred. No. 0.88;
  Matches   41;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy         68 TACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTA 127
              ||  ||||| |||||  | | |||| ||| | | ||| |||  || ||| ||  || ||
Db        202 TAAGAATGGTAGCAAGAGGAATGAGTATTCTACGCATCACACCATCACTGCTGTTCACTA 143


RESULT 9
US-08-793-273C-1
; Sequence 1, Application US/08793273C
; Patent No. 6482410
; GENERAL INFORMATION:
;  APPLICANT: Crossin, Kathryn L.
;  APPLICANT:  Phillips, Greg
;  APPLICANT:  Prieto, Anne L.
;  TITLE OF INVENTION: CYTOTACTIN DERIVATIVES THAT STIMULATE ATTACHMENT AND
;  TITLE OF INVENTION:  NEURITE OUTGROWTH, AND METHODS OF MAKING SAME
;  FILE REFERENCE: BEC0022S
;  CURRENT APPLICATION NUMBER: US/08/793,273C
;  CURRENT FILING DATE:  1997-05-22
;  PRIOR APPLICATION NUMBER: PCT/US95/11684
;  PRIOR FILING DATE: 1995-09-14
;  PRIOR APPLICATION NUMBER: 08/308,359
;  PRIOR FILING DATE: 1994-09-16
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;    LENGTH: 7286
;    TYPE: DNA
;    ORGANISM: Homo Sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (55)..(6654)
US-08-793-273C-1

  Query Match             12.4%;  Score 29.2;  DB 4;  Length 7286;
  Best Local Similarity   50.0%;  Pred. No. 4.6;
  Matches   73;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              | || | |  ||| |     | || |||| |        ||| ||  | |  |||||||
Db       2789 GGAGGAATGGCAAGGCAGCTATTGACAGTTACAGAATTAAGTATGCCCCCATCTCTGGAG 2848

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              || |  | | || || ||  | ||| ||    ||  |      |      ||   |||
Db       2849 GGGACCACGCTGAGGTTGATGTTCCAAAGAGCCAACAAGCCACAACCAAAACCACACTCA 2908

Qy        181 CTGGAGTGATGCCTAATATTCAGTAT 206
              | ||  ||| |||    | | |||
Db       2909 CAGGTCTGAGGCCGGGAACTGAATAT 2934


RESULT 10
PCT-US95-11684-1
; Sequence 1, Application PC/TUS9511684
;  GENERAL INFORMATION:
;    APPLICANT:  THE SCRIPPS RESEARCH INSTITUTE
;    TITLE OF INVENTION:  CYTOTACTIN DERIVATIVES THAT STIMULATE
;    TITLE OF INVENTION:  ATTACHMENT AND NEURITE OUTGROWTH, AND METHODS OF MAKING
;    TITLE OF INVENTION:  AND USING SAME
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 North Torrey Pines Road, TPC 8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/11684
;      FILING DATE:  14-SEP-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,359
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Logan, April C.
;      REGISTRATION NUMBER:  33,950
;      REFERENCE/DOCKET NUMBER:  BEC0019P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7286 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..6654
;      OTHER INFORMATION:  /product= "cytotactin"
PCT-US95-11684-1

  Query Match             12.4%;  Score 29.2;  DB 5;  Length 7286;
  Best Local Similarity   50.0%;  Pred. No. 4.6;
  Matches   73;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;
```

```
Qy          61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
               | || | |  ||| |     | || |||| |        ||| ||  | |  ||||||| |
Db        2789 GGAGGAATGGCAAGGCAGCTATTGACAGTTACAGAATTAAGTATGCCCCCATCTCTGGAG 2848

Qy         121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
               || |  | | || || ||  | ||| ||    ||  |       |      ||   |||
Db        2849 GGGACCACGCTGAGGTTGATGTTCCAAAGAGCCAACAAGCCACAACCAAAACCACACTCA 2908

Qy         181 CTGGAGTGATGCCTAATATTCAGTAT 206
               | ||  ||| |||    | | | |||
Db        2909 CAGGTCTGAGGCCGGGAACTGAATAT 2934


RESULT 11
JS-08-958-201-1/c
; Sequence 1, Application US/08958201
; Patent No. 5977319
;  GENERAL INFORMATION:
;    APPLICANT:  Pope, Anthony R
;    APPLICANT:  Pritchard, Kevin
;    APPLICANT:  Williams, Andrew J
;    APPLICANT:  Johnson, Kevin S
;    TITLE OF INVENTION:  Specific binding members for estradiol;
;    TITLE OF INVENTION:  materials and methods
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall O'Toole Gerstein Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/958,201
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/028,897
;      FILING DATE:  21-OCT-1996
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  354 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  D12
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..354
JS-08-958-201-1

  Query Match             12.3%;  Score 29;  DB 2;  Length 354;
  Best Local Similarity   54.1%;  Pred. No. 1.1;
  Matches   59;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          78 AGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCT 137
               | |||   || | |||  ||| | ||||||||||    ||| |      | |   |
Db         318 ATCAAAACCATTTAGTACTCCTAAAATCGCACAGTAATATATGGCCGTGTCCTCGGCTCG 259

Qy         138 GCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTACCTGGAG 186
               |||    | || ||||  || |  |||||   |    | |||||||
Db         258 CAGGCTGTTCATTTGCAGATACACCGTGTTCTTGGAATTGTCCCTGGAG 210


RESULT 12
JS-09-831-206-1/c
; Sequence 1, Application US/09831206
; Patent No. 6573070
; GENERAL INFORMATION:
;  APPLICANT: MacNeil, Douglas J.
;  APPLICANT:  Weinberg, David H.
;  APPLICANT:  Van der Ploeg, Leonardus H. T.
;  TITLE OF INVENTION: DNA MOLECULES ENCODING THE MELANOCORTIN
;  TITLE OF INVENTION:  4 RECEPTOR PROTEIN FROM RHESUS MONKEY
;  FILE REFERENCE: 20190P
;  CURRENT APPLICATION NUMBER: US/09/831,206
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: PCT/US99/25767
;  PRIOR FILING DATE: 1999-11-05
;  PRIOR APPLICATION NUMBER: 60/107,721
;  PRIOR FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1030
;   TYPE: DNA
;   ORGANISM: rhesus monkey (Macaca mulatta)
US-09-831-206-1

  Query Match             12.3%;  Score 28.8;  DB 4;  Length 1030;
  Best Local Similarity   65.6%;  Pred. No. 2.2;
  Matches   42;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          93 TCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTGCCGCTCCCAACTTA 152
               | |||||||   | ||| | |  || || ||||| ||| |  |||| | |||  |  |
Db         115 TTTTCCAAGGGACTCACTGGCATTGCTGTGCAGTCTGTGGCTGCTGCGGTTCCAGAGGTG 56

Qy         153 CAGA 156
               ||||
Db          55 CAGA 52


RESULT 13
US-09-238-303-7
; Sequence 7, Application US/09238303B
; Patent No. 6284253
; GENERAL INFORMATION:
;  APPLICANT: Barr, Margaret C.
;  TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequence
;  FILE REFERENCE: 18617.0059
;  CURRENT APPLICATION NUMBER: US/09/238,303B
;  CURRENT FILING DATE: 1999-01-28
;  EARLIER APPLICATION NUMBER: US 60/072,927
;  EARLIER FILING DATE: 1998-01-29
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 7
;   LENGTH: 9751
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of
;   OTHER INFORMATION: a Pallas's cat feline immunodeficiency virus
US-09-238-303-7

  Query Match             12.3%;  Score 28.8;  DB 3;  Length 9751;
  Best Local Similarity   56.2%;  Pred. No. 7.4;
  Matches   54;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy         121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
               ||||| | |     ||||| || || | |  |||    |||||| ||| |||| | ||
Db        2045 GGCAGAAGGGGAGAGCTGCTGCCCCTATCAAACAAGTGCAGCAATTTCAAACAGCAGTAT 2104

Qy         181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCC 216
               |   |     |  | |    || | |||| |||| |
Db        2105 CAACAACTCAGAATCAGCAACAATGTCAATTAATAC 2140


RESULT 14
US-09-946-239-7
; Sequence 7, Application US/09946239
```

```
; Patent No. 6579527
; GENERAL INFORMATION:
;  APPLICANT: Barr, Margaret C.
;  TITLE OF INVENTION: No. 6579527el Feline Immunodeficiency Virus Nucleotide and
;  TITLE OF INVENTION:  Polypeptide Sequences
;  FILE REFERENCE: 18617.0059
;  CURRENT APPLICATION NUMBER: US/09/946,239
;  CURRENT FILING DATE:  2001-09-04
;  PRIOR APPLICATION NUMBER: US 09/238,303, US 60/072,927
;  PRIOR FILING DATE: 1999-01-28, 1998-01-29
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 7
;   LENGTH: 9751
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of
;   OTHER INFORMATION: a Pallas's cat feline immunodeficiency virus
US-09-946-239-7

  Query Match             12.3%;  Score 28.8;  DB 4;  Length 9751;
  Best Local Similarity   56.2%;  Pred. No. 7.4;
  Matches   54;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||| | |     ||||| || || | |  |||     |||||| |||  |||| | ||
Db       2045 GGCAGAAGGGGAGAGCTGCTGCCCCTATCAAACAAGTGCAGCAATTTCAAACAGCAGTAT 2104

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCC 216
              |   |     | | |    || | |||| |||| |
Db       2105 CAACAACTCAGAATCAGCAACAATGTCAATTAATAC 2140


RESULT 15
US-09-634-238-409
; Sequence 409, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
;  APPLICANT:  Glenn, Matthew
;  APPLICANT:  Havukkala, Ilkka J.
;  APPLICANT:  Bloksberg, Leonard, N.
;  APPLICANT:  Lubbers, Mark W.
;  APPLICANT:  Dekker, James
;  APPLICANT:  Christensson, Anna C.
;  APPLICANT:  Holland, Ross
;  APPLICANT:  O'Toole, Paul W.
;  APPLICANT:  Reid, Julian R.
;  APPLICANT:  Coolbear, Timothy
;  TITLE OF INVENTION:  Polynucleotides, materials incorporating
;  TITLE OF INVENTION:  them and methods for using them.
;  FILE REFERENCE:  11000.1043U1
;  CURRENT APPLICATION NUMBER: US/09/634,238
;  CURRENT FILING DATE:  2000-08-08
;  NUMBER OF SEQ ID NOS:  422
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 409
;   LENGTH: 3269
;   TYPE: DNA
;   ORGANISM: Lactobacillus rhamnosus
US-09-634-238-409

  Query Match             12.1%;  Score 28.4;  DB 4;  Length 3269;
  Best Local Similarity   50.7%;  Pred. No. 5.7;
  Matches   68;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy         20 TTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAG 79
              | ||| | |       |  | |||| ||||||  || |     |||| |  ||
Db       1343 TGCCTTCGCCTCAAGTGGCGCTACTTCAAAGTGACTGGTAATCAGCAATTGCATCACATT 1402

Qy         80 CAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTGC 139
                |   |     ||   || |   ||||| | ||| | |||| |    ||||   |||||
Db       1403 GGCTTCAACATCATCGAGCATGTCGCGCACCGGCTCCGTTGGGTAAACTGTTACCGCTGC 1462

Qy        140 CGCTCCCAACTTAC 153
              | |     ||||||
Db       1463 CTCCATGGACTTAC 1476


Search completed: February 18, 2004, 16:10:41
Job time : 63.179 secs
```

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 14:50:26 ; Search time 275.233 Seconds
(without alignments)
3145.162 Million cell updates/sec

Title: US-10-026-341A-1
Perfect score: 235
Sequence: 1 atggctggcagatcatctct.....ccacagttccagaccgttga 235

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	235	100.0	235	14	US-10-026-341A-1	Sequence 1, Appli
2	235	100.0	3289	13	US-10-117-722-436	Sequence 436, App
3	235	100.0	3289	15	US-10-037-270-436	Sequence 436, App
4	69.8	29.7	570	13	US-09-814-353-18291	Sequence 18291, A
5	65.4	27.8	682	13	US-09-814-353-21704	Sequence 21704, A
6	50	21.3	411	13	US-09-814-353-5620	Sequence 5620, Ap
7	50	21.3	411	13	US-09-814-353-11907	Sequence 11907, A
c 8	33	14.0	2732	15	US-10-102-524-1754	Sequence 1754, Ap
c 9	33	14.0	2865	13	US-09-873-367C-135	Sequence 135, App
10	31.6	13.4	147309	9	US-09-742-312-3	Sequence 3, Appli
11	31.6	13.4	147309	13	US-10-436-185-3	Sequence 3, Appli
12	31.2	13.3	4452	12	US-10-085-117-74	Sequence 74, Appl
c 13	31.2	13.3	99934	12	US-10-085-117-73	Sequence 73, Appl
14	30.8	13.1	3231	13	US-10-056-454A-18	Sequence 18, Appl
c 15	30.8	13.1	130001	12	US-10-199-676-23	Sequence 23, Appl
c 16	30.2	12.9	359	9	US-09-922-217-801	Sequence 801, App
c 17	30.2	12.9	359	10	US-09-833-263-801	Sequence 801, App
c 18	30.2	12.9	359	10	US-09-920-300A-1213	Sequence 1213, Ap
c 19	30.2	12.9	359	13	US-10-099-926-1213	Sequence 1213, Ap
c 20	30.2	12.9	359	14	US-10-033-528-1213	Sequence 1213, Ap
c 21	30.2	12.9	359	14	US-10-025-380-801	Sequence 801, App
22	30.2	12.9	1076	15	US-10-097-340-325	Sequence 325, App
23	30.2	12.9	1076	15	US-10-205-823-424	Sequence 424, App
24	30.2	12.9	1278	12	US-10-295-027-603	Sequence 603, App
25	30.2	12.9	1288	16	US-10-156-136-4	Sequence 4, Appli
c 26	30.2	12.9	1288	16	US-10-156-136-50	Sequence 50, Appl
27	30.2	12.9	1289	9	US-09-759-143-111	Sequence 111, App
28	30.2	12.9	1289	9	US-09-780-669-111	Sequence 111, App
29	30.2	12.9	1289	9	US-09-030-606-111	Sequence 111, App
30	30.2	12.9	1289	9	US-09-822-827-111	Sequence 111, App
31	30.2	12.9	1289	9	US-09-115-453-111	Sequence 111, App
32	30.2	12.9	1289	10	US-09-232-880-111	Sequence 111, App
33	30.2	12.9	1289	10	US-09-895-793-111	Sequence 111, App
34	30.2	12.9	1289	10	US-09-895-814-111	Sequence 111, App
35	30.2	12.9	1289	13	US-10-144-678A-111	Sequence 111, App
36	30.2	12.9	1289	13	US-10-294-025-111	Sequence 111, App
37	30.2	12.9	1289	14	US-10-012-896-111	Sequence 111, App
38	30.2	12.9	1289	15	US-10-010-940-111	Sequence 111, App
39	30.2	12.9	1324	10	US-09-981-876-82	Sequence 82, Appl
40	30.2	12.9	1324	11	US-09-148-545-82	Sequence 82, Appl
41	30.2	12.9	1346	12	US-10-264-237-84	Sequence 84, Appl
42	30.2	12.9	1376	10	US-09-981-876-14	Sequence 14, Appl
43	30.2	12.9	1376	11	US-09-148-545-14	Sequence 14, Appl
44	30.2	12.9	1456	10	US-09-981-353-133	Sequence 133, App
45	30.2	12.9	1494	10	US-09-981-876-83	Sequence 83, Appl

ALIGNMENTS

```
RESULT 1
US-10-026-341A-1
; Sequence 1, Application US/10026341A
; Publication No. US20020137715A1
; GENERAL INFORMATION:
;  APPLICANT: Alain Mauviel
;  TITLE OF INVENTION: Blocking Sp1 Transcription Factor
;  TITLE OF INVENTION:  Broadly Inhibits Extracellular Matrix Gene Expression In
;  TITLE OF INVENTION:  Vitro and In Vivo: Implications for the Treatment of Tissue
;  TITLE OF INVENTION:  Fibrosis
;  FILE REFERENCE: MAU01.NP001
;  CURRENT APPLICATION NUMBER: US/10/026,341A
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/259,585
;  PRIOR FILING DATE: 2001-01-03
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 235
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic oligonucleotide
US-10-026-341A-1

  Query Match             100.0%;  Score 235;  DB 14;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-72;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235


RESULT 2
US-10-117-722-436
; Sequence 436, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2BCIP
;  CURRENT APPLICATION NUMBER: US/10/117,722
;  CURRENT FILING DATE:  2002-04-04
;  PRIOR APPLICATION NUMBER: 09/620,312
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 436
;   LENGTH: 3289
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(2558)
US-10-117-722-436

  Query Match             100.0%;  Score 235;  DB 13;  Length 3289;
  Best Local Similarity   100.0%;  Pred. No. 1e-71;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 573

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 633

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 693

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 748


RESULT 3
US-10-037-270-436
; Sequence 436, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Tillinghast, John
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/10/037,270
;  CURRENT FILING DATE:  2002-01-04
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 436
;   LENGTH: 3289
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(2558)
US-10-037-270-436

  Query Match             100.0%;  Score 235;  DB 15;  Length 3289;
  Best Local Similarity   100.0%;  Pred. No. 1e-71;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 573

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 633

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 693

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 748


RESULT 4
US-09-814-353-18291
; Sequence 18291, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
```

```
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18291
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-18291

  Query Match             29.7%;  Score 69.8;  DB 13;  Length 570;
  Best Local Similarity   91.4%;  Pred. No. 5.3e-14;
  Matches   74;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        430 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAATG 489

Qy         61 GCAGCAGTACCAATGGCAGCA 81
              |||||||||||  | |  ||
Db        490 GCAGCAGTACCTCGGCCGCCA 510


RESULT 5
US-09-814-353-21704
; Sequence 21704, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21704
;   LENGTH: 682
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 682
;   OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21704

  Query Match             27.8%;  Score 65.4;  DB 13;  Length 682;
  Best Local Similarity   98.5%;  Pred. No. 2e-12;
  Matches   66;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        615 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAATG 674

Qy         61 GCAGCAG 67
              |||||||
Db        675 GCAGCAG 681


RESULT 6
US-09-814-353-5620
; Sequence 5620, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5620
;   LENGTH: 411
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-5620

  Query Match             21.3%;  Score 50;  DB 13;  Length 411;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-07;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAG 50
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAG 411


RESULT 7
US-09-814-353-11907
; Sequence 11907, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11907
;   LENGTH: 411
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-11907

  Query Match             21.3%;  Score 50;  DB 13;  Length 411;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-07;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAG 50
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAG 411


RESULT 8
US-10-102-524-1754/c
; Sequence 1754, Application US/10102524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
;  APPLICANT: Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Gaiger, Alexander
;  APPLICANT:  Gordon, Brian
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF KIDNEY CANCER
;  FILE REFERENCE: 210121.572
;  CURRENT APPLICATION NUMBER: US/10/102,524
;  CURRENT FILING DATE:  2002-03-19
;  NUMBER OF SEQ ID NOS: 1863
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1754
;   LENGTH: 2732
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-102-524-1754

  Query Match             14.0%;  Score 33;  DB 15;  Length 2732;
  Best Local Similarity   57.1%;  Pred. No. 0.85;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         21 TCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGC 80
              ||||| || ||| |  |  |  || | | ||     ||  ||| |||  ||||  |||||
Db        133 TCCTCGTCCGGGCCCTCGGCACCCGCCACGGCCAGCAGCTGCACCAGCCCCAAGAGCAGC 74

Qy         81 AATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAG 125
              |   ||||  || ||| |  |   | ||  ||  | || | | ||
Db         73 AGGAGCAGGAAGGCTTTCCGGGGCCTCATGGTAGCGGGGGCGGAG 29


RESULT 9
US-09-873-367C-135/c
; Sequence 135, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  APPLICANT:  Soppet, Daniel
;  APPLICANT:  Endress, Gregory
;  APPLICANT:  Augustus, Meena
;  APPLICANT:  Ebner, Reinhard
;  APPLICANT:  Carter, Kenneth
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
;  TITLE OF INVENTION:  Signature Gene Sets
;  FILE REFERENCE: 689290-64
;  CURRENT APPLICATION NUMBER: US/09/873,367C
;  CURRENT FILING DATE:  2003-04-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,891
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/236,842
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: U.S. 60/244,867
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: U.S. 60/245,084
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 1067
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 135
;   LENGTH: 2865
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-873-367C-135

  Query Match             14.0%;  Score 33;  DB 13;  Length 2865;
  Best Local Similarity   57.1%;  Pred. No. 0.87;
  Matches   60;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         21 TCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGC 80
              ||||| || ||| |  |  |  || | | ||     ||  ||| |||  ||||  |||||
Db        134 TCCTCGTCCGGGCCCTCGGCACCCGCCACGGCCAGCAGCTGCACCAGCCCCAAGAGCAGC 75

Qy         81 AATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAG 125
              |   ||||  || ||| |  |   | ||  ||  | || | | ||
Db         74 AGGAGCAGGAAGGCTTTCCGGGGCCTCATGGTAGCGGGGGCGGAG 30


RESULT 10
US-09-742-312-3
; Sequence 3, Application US/09742312
; Patent No. US20020045166A1
; GENERAL INFORMATION:
;  APPLICANT: CHANDRAMOULISARAN, Ishwar et al
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000838
;  CURRENT APPLICATION NUMBER: US/09/742,312
;  CURRENT FILING DATE:  2000-12-22
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 147309
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(147309)
;   OTHER INFORMATION: n = A,T,C or G
US-09-742-312-3

  Query Match             13.4%;  Score 31.6;  DB 9;  Length 147309;
  Best Local Similarity   58.5%;  Pred. No. 14;
  Matches   55;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy        130 TTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTACCTGGAGTGA 189
              |||  |||||| ||||||  |       ||  |||   |||||  |  | |  | || ||||||
Db      97837 TTGCTGCTGCCCCCTCCCACCCCCATGACTCAGTCTGTTCTCAATGCAGGAGCTAGAGTGA 97896

Qy        190 TGCCTAATATTCAGTATCAAGTAATCCCACAGTT 223
              | | |   |  |   ||  |||  || ||  |||
Db      97897 TCCATTTAAAACGTAATTGAGTCTTCTCATTGTT 97930


RESULT 11
US-10-436-185-3
; Sequence 3, Application US/10436185
; Publication No. US20030180887A1
; GENERAL INFORMATION:
;  APPLICANT: CHANDRAMOULISARAN, Ishwar et al
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000838CON
;  CURRENT APPLICATION NUMBER: US/10/436,185
;  CURRENT FILING DATE:  2003-05-13
```

```
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;    LENGTH: 147309
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(147309)
;    OTHER INFORMATION: n = A,T,C or G
US-10-436-185-3

  Query Match             13.4%;  Score 31.6;  DB 13;  Length 147309;
  Best Local Similarity   58.5%;  Pred. No. 14;
  Matches   55;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy        130 TTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTACCTGGAGTGA 189
              ||| |||||| |||||| |      || |||   ||||| | | |  | || ||||||
Db      97837 TTGCTGCTGCCCCTCCCACCCCCATGACTCAGTCTGTTCTCAATGCAGGAGCTAGAGTGA 97896

Qy        190 TGCCTAATATTCAGTATCAAGTAATCCCACAGTT 223
              | | |    |  |    ||  |||  || ||  |||
Db      97897 TCCATTTAAAACGTAATTGAGTCTTCTCATTGTT 97930


RESULT 12
US-10-085-117-74
; Sequence 74, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: 529452000121
;  CURRENT APPLICATION NUMBER: US/10/085,117
;  CURRENT FILING DATE:  2002-02-27
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 361
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;    LENGTH: 4452
;    TYPE: DNA
;    ORGANISM: Mus musculus
JS-10-085-117-74

  Query Match             13.3%;  Score 31.2;  DB 12;  Length 4452;
  Best Local Similarity   58.7%;  Pred. No. 4.5;
  Matches   54;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         45 TCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAAT 104
              | | |||||| | ||| |  |  | |||| |  || || | ||   ||  || | ||| |
Db       2842 TGATAGGAACTGTGTGTCCACCTTTCCAAAGAAAGGAAAGACAAAAAGAGTTTCCAGATT 2901

Qy        105 CGCACAGTCTCTGGTGGGCAGTATGTTGTGGC 136
              || |||   ||| | ||   | | || || |
Db       2902 GGCCCAGCTGCTGATAGGTGATTTTTTTTGCC 2933


RESULT 13
JS-10-085-117-73/c
; Sequence 73, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: 529452000121
;  CURRENT APPLICATION NUMBER: US/10/085,117
;  CURRENT FILING DATE:  2002-02-27
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 361
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 73
;    LENGTH: 99934
;    TYPE: DNA
;    ORGANISM: Mus musculus
;    FEATURE:
;    NAME/KEY: variation
;    LOCATION: (1)...(99934)
;    OTHER INFORMATION: n = any nucleotide
US-10-085-117-73

  Query Match             13.3%;  Score 31.2;  DB 12;  Length 99934;
  Best Local Similarity   58.7%;  Pred. No. 16;
  Matches   54;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         45 TCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAAT 104
              | | |||||| | ||| |  |  | |||| |  || || | ||   ||  || | ||| |
Db      11611 TGATAGGAACTGTGTGTCCACCTTTCCAAAGAAAGGAAAGACAAAAAGAGTTTCCAGATT 11552

Qy        105 CGCACAGTCTCTGGTGGGCAGTATGTTGTGGC 136
              || |||   ||| | ||   | | || || |
Db      11551 GGCCCAGCTGCTGATAGGTGATTTTTTTTGCC 11520


RESULT 14
US-10-056-454A-18
; Sequence 18, Application US/10056454A
; Publication No. US20030166919A1
;   GENERAL INFORMATION:
;        APPLICANT: National Starch and Chemical Investment Holding Corporation
;        TITLE OF INVENTION: Improvements in or Relating to Plant Starch Composition
;        NUMBER OF SEQUENCES: 20
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: National Starch and Chemical Investment Holding Corporation
;             STREET: 1000 Uniqema Blvd.
;             CITY: Newcastle
;             STATE: Delaware
;             COUNTRY: United States of America
;             ZIP: 19720
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/056,454A
;             FILING DATE: 25-Jun-2002
;   INFORMATION FOR SEQ ID NO: 18:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 3231 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-056-454A-18

  Query Match             13.1%;  Score 30.8;  DB 13;  Length 3231;
  Best Local Similarity   54.4%;  Pred. No. 5.4;
  Matches   62;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         41 TACCTCAAAGGAACAGAGTGGCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAA 100
              |  | ||| ||||||   | ||   | ||  | ||  | | ||| |  |||| | ||| |
Db        536 TTCAACAATGGAACACGCTAGCCAGATTAAAACTGAGAACGATGACGTTGAGCCGTCAAG 595

Qy        101 GAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTGCCGCTCCCAACTTACA 154
                 |||  ||||    || ||   |||  | | | | ||| |  | |   |  ||||
Db        596 TGATCTTACAGGAAGTGTTGAAGAGTTGGATTTTGCTTCATCACTACAACTACA 649


RESULT 15
US-10-199-676-23/c
; Sequence 23, Application US/10199676
; Publication No. US20040014051A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Vickie L. Brown-Driver
;  APPLICANT:  Kenneth W. Dobie
;  TITLE OF INVENTION: ANTISENSE MODULATION OF BREAST CANCER-1 EXPRESSION
;  FILE REFERENCE: PTS-0017
;  CURRENT APPLICATION NUMBER: US/10/199,676
;  CURRENT FILING DATE:  2002-07-18
;  NUMBER OF SEQ ID NOS: 84
; SEQ ID NO 23
;   LENGTH: 130001
;   TYPE: DNA
;   ORGANISM: H. sapiens
;   FEATURE:
US-10-199-676-23

  Query Match             13.1%;  Score 30.8;  DB 12;  Length 130001;
  Best Local Similarity   48.8%;  Pred. No. 25;
  Matches   83;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy          3 GGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTGGC 62
              ||| ||  | ||      ||||   || | | | | ||| |||||      || ||||
Db     128664 GGCAGGATGGTCTCGAACTCCTGACCTCGTGATCCGCCTGCCTCAGCCTCCCAAAGTGCT 128605

Qy         63 AGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGG 122
               | | |||    |  ||| |  |     ||  |  | |   |     ||   | ||
Db     128604 GGGATTACAGGCGTGAGCCACCGTGCCCAGCATGGCTAATTTTTTGTAGAGACAGGGTTT 128545

Qy        123 CAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGAC 172
              ||| ||||||     || | || | |  | | || |   || | |||| |
Db     128544 CAGCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCACATGATCTGCC 128495


Search completed: February 18, 2004, 16:15:54
Job time : 277.233 secs
```

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 12:26:35 ; Search time 2369.2 Seconds
(without alignments)
2410.749 Million cell updates/sec

Title: US-10-026-341A-1
Perfect score: 235
Sequence: 1 atggctggcagatcatctct.....ccacagttccagaccgttga 235

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	235	100.0	755	12	BI559366	BI559366 603253037
2	233.4	99.3	645	9	AW609227	AW609227 RC3-ST018
3	233.4	99.3	949	13	BQ890167	BQ890167 AGENCOURT
4	222.4	94.6	1150	13	BQ072120	BQ072120 AGENCOURT
5	192.6	82.0	546	12	BG803579	BG803579 0233-17 M
6	192.6	82.0	617	9	AW822630	AW822630 uq13d03.y
7	186.8	79.5	590	10	BE937236	BE937236 RC3-ST018
8	183.2	78.0	653	14	BY738990	BY738990 BY738990
9	183	77.9	586	13	BU702181	BU702181 UI-M-FI0-
10	157.4	67.0	416	9	AW445976	AW445976 83265 MAR
11	147.2	62.6	404	14	CB811553	CB811553 AMGNNUC:S
12	75.4	32.1	681	12	BM427106	BM427106 pgf2n.pk0
c 13	70	29.8	504	12	BG997913	BG997913 PM1-HT129
14	67.2	28.6	880	13	BU906803	BU906803 AGENCOURT
15	67.2	28.6	928	14	CA787807	CA787807 AGENCOURT
16	61.8	26.3	922	14	CA987845	CA987845 AGENCOURT
17	61.4	26.1	670	9	AL647036	AL647036 AL647036
18	61.4	26.1	671	13	BQ397449	BQ397449 NISC_ng27
19	58	24.7	450	10	BF875151	BF875151 QV3-ET010
20	57	24.3	382	10	BF510251	BF510251 UI-H-BI4-
21	50.4	21.4	629	14	CD279637	CD279637 G44222.67
22	50.4	21.4	755	14	CA470150	CA470150 AGENCOURT
23	50.4	21.4	832	14	CA476365	CA476365 AGENCOURT
24	48.8	20.8	473	14	CD280695	CD280695 G43520.14
25	48.8	20.8	902	14	CA469702	CA469702 AGENCOURT
26	48.4	20.6	740	14	CB940752	CB940752 IpCGJx14_
c 27	40	17.0	257	9	AW178522	AW178522 CM0-HT011_
28	39.8	16.9	502	10	BE696452	BE696452 MR1-CT039
29	39.4	16.8	622	14	CD282914	CD282914 G39156.92
30	38.4	16.3	954	13	BU236388	BU236388 603407867
31	36.2	15.4	810	12	BI837393	BI837393 603086375
32	35.8	15.2	527	14	CD036456	CD036456 mgmt020xL
33	35.8	15.2	642	14	CA622180	CA622180 wl1n.pk00
c 34	35.6	15.1	744	10	BG480269	BG480269 602529310
c 35	34.6	14.7	675	9	AW027138	AW027138 wv65h05.x
36	34.6	14.7	804	12	BI837784	BI837784 603086475
c 37	34.6	14.7	1201	9	AL523072	AL523072 AL523072
c 38	34.6	14.7	1201	9	AL529564	AL529564 AL529564
39	34.4	14.6	874	14	CD302448	CD302448 AGENCOURT
40	34.2	14.6	1201	9	AL523073	AL523073 AL523073
41	33.8	14.4	448	29	BZ867528	BZ867528 CH240_274
42	33.8	14.4	673	29	AG096615	AG096615 Pan trogl
c 43	33.8	14.4	1201	13	BX394709	BX394709 BX394709
44	33.6	14.3	273	9	AV328003	AV328003 AV328003
c 45	33.4	14.2	859	14	CB559451	CB559451 AGENCOURT

ALIGNMENTS

RESULT 1
BI559366
LOCUS BI559366 755 bp mRNA linear EST 05-SEP-2001
DEFINITION 603253037F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5295465 5', mRNA sequence.
ACCESSION BI559366
VERSION BI559366.1 GI:15446680
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 755)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11747 row: j column: 10

```
          High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1. .755
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5295465"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_97"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      199 a    203 c    207 g    146 t
ORIGIN

  Query Match             100.0%;  Score 235;  DB 12;  Length 755;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-58;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 412

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        413 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 472

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        473 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 532

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 587


RESULT 2
AW609227
LOCUS       AW609227                645 bp    mRNA    linear   EST 23-MAR-2000
DEFINITION  RC3-ST0186-300100-017-f03 ST0186 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW609227
VERSION     AW609227.1  GI:7313968
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC3&t2=RC3-ST0186-
            300100-017-f03&t3=2000-01-30&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 644.
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="ST0186"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      158 a    152 c    189 g    146 t
ORIGIN

  Query Match             99.3%;  Score 233.4;  DB 9;  Length 645;
  Best Local Similarity   99.6%;  Pred. No. 2.4e-57;
  Matches  234;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 393

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 453

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 513

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 CTGGAGTGATGCCTAATATTCAGTATCAAGTACTCCCACAGTTCCAGACCGTTGA 568


RESULT 3
BQ890167
LOCUS       BQ890167                949 bp    mRNA    linear   EST 16-AUG-2002
DEFINITION  AGENCOURT_8730196 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6339589
            5', mRNA sequence.
ACCESSION   BQ890167
VERSION     BQ890167.1  GI:22282181
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 949)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2534  row: k  column: 14
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .949
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6339589"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_47"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
```

EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 239 a 269 c 258 g 183 t
ORIGIN

Query Match 99.3%; Score 233.4; DB 13; Length 949;
Best Local Similarity 99.6%; Pred. No. 2.9e-57;
Matches 234; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        411 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 470

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        471 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 530

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        531 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAAACAGCAAGTTCTGACAGGACTAC 590

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        591 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 645
```

RESULT 4
BQ072120
LOCUS BQ072120 1150 bp mRNA linear EST 02-APR-2002
DEFINITION AGENCOURT_6859825 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5928633 5', mRNA sequence.
ACCESSION BQ072120
VERSION BQ072120.1 GI:19901166
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1150)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2105 row: h column: 10
High quality sequence stop: 571.
FEATURES Location/Qualifiers
source 1. .1150
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5928633"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_47"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 281 a 344 c 290 g 235 t
ORIGIN

Query Match 94.6%; Score 222.4; DB 13; Length 1150;
Best Local Similarity 99.2%; Pred. No. 5.2e-54;
Matches 234; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        411 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 470

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        471 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 530

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        531 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 590

Qy        181 CTGG-AGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||| ||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        591 CTGGAAGTGATGCCTAATATTCAGTATCCAGTAATCCCACAGTTCCAGACCGTTGA 646
```

RESULT 5
BG803579
LOCUS BG803579 546 bp mRNA linear EST 20-DEC-2001
DEFINITION 0233-17 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA, mRNA sequence.
ACCESSION BG803579
VERSION BG803579.1 GI:17950485
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 546)
AUTHORS Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W., White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
TITLE Gene expression in the developing mouse retina by EST sequencing and microarray analysis
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE 21671825
PUBMED 11812828
COMMENT Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.
FEATURES Location/Qualifiers
source 1. .546
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT 139 a 154 c 129 g 124 t
ORIGIN

Query Match 82.0%; Score 192.6; DB 12; Length 546;
Best Local Similarity 92.3%; Pred. No. 1.7e-45;
Matches 217; Conservative 0; Mismatches 9; Indels 9; Gaps 1;

```
Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 77
```

```
Qy          61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
               ||| ||||||||||||||||          |||||||||||||| ||||||||||||||||
Db          78 GCAACAGTACCAATGGCAGC---------GAGTCTTCCAAGAACCGCACAGTCTCTGGTG 128

Qy         121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
               |||||||||||||||||| | ||||||||||||||||||||||||||||||| || |
Db         129 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTCC 188

Qy         181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
               ||||||| |||||||||||||||||||||||||||||||||||||| |||||
Db         189 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 243
```

RESULT 6
AW822630

```
LOCUS       AW822630                617 bp    mRNA    linear   EST 17-MAY-2000
DEFINITION  uq13d03.y1 Ren Stubbs mouse thymus Mus musculus cDNA clone
            IMAGE:2802341 5' similar to gb:J03133 TRANSCRIPTION FACTOR SP1
            (HUMAN);, mRNA sequence.
ACCESSION   AW822630
VERSION     AW822630.1  GI:7915647
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished
COMMENT     Other_ESTs: uq13d03.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1041865
            Seq primer: Primer name ambiguous
            High quality sequence stop: 476.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2802341"
                     /sex="mixed"
                     /dev_stage="3 weeks"
                     /lab_host="DH10B"
                     /clone_lib="Ren Stubbs mouse thymus"
                     /note="Organ: thymus; Vector: pT7T3D-Pac; Site_1: NotI;
                     Site_2: PacI; 1st strand cDNA was primed with an oligo(dT)
                     primer; double-stranded cDNA was ligated using 5' linker
                     ggccgctat and 3' linker aactggaagcttaatt. Library is
                     size-selected >2.5 kb and average insert size is 3.5 kb.
                     Clones were arrayed from primary plating; non-amplified.
                     Library constructed by X. Ren and L. Stubbs (Lawrence
                     Livermore National Laboratory and DOE Joint Genome
                     Institute, 7000 East Ave, L-453, Livermore, CA 94550)."
BASE COUNT      154 a    173 c    169 g    121 t
ORIGIN
```

```
  Query Match             82.0%;  Score 192.6;  DB 9;  Length 617;
  Best Local Similarity   92.3%;  Pred. No. 1.8e-45;
  Matches  217;  Conservative    0;  Mismatches    9;  Indels    9;  Gaps    1;
```

```
Qy           1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         356 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 415

Qy          61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
               ||| ||||||||||||||||          |||||||||||||| ||||||||||||||||
Db         416 GCAACAGTACCAATGGCAGC---------GAGTCTTCCAAGAACCGCACAGTCTCTGGTG 466

Qy         121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
               |||||||||||||||||| | ||||||||||||||||||||||||||||||| || |
Db         467 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTCC 526

Qy         181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
               ||||||| |||||||||||||||||||||||||||||||||||||| |||||
Db         527 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 581
```

RESULT 7
BE937236

```
LOCUS       BE937236                590 bp    mRNA    linear   EST 02-OCT-2000
DEFINITION  RC3-ST0186-290800-013-d04 ST0186 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE937236
VERSION     BE937236.1  GI:10463460
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC3-ST0186-290
            800-013-d04&t3=2000-08-29&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 2
            High quality sequence stop: 506.
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="ST0186"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      151 a    133 c    172 g    134 t
ORIGIN
```

```
  Query Match             79.5%;  Score 186.8;  DB 10;  Length 590;
```

```
  Best Local Similarity   97.5%;  Pred. No. 8.8e-44;
  Matches  232;  Conservative    0;  Mismatches    2;  Indels    4;  Gaps    4;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 411

Qy         61 GCAGCAGTACCAATGGCA-GCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGT 119
              |||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db        412 GCAGCAGTACCAGTGGCACGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGT 471

Qy        120 -GGGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACT 178
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        472 GGGGCAGTATGTTGTGGCTG-CGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACT 530

Qy        179 ACCTGGAGTGATGCCTAATATTCAGTATCAAGTAA-TCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||| ||| |||||||||||||||||
Db        531 ACCTGGAGTGATGCCTAATATTCAGTATCAAGTAACTCCAACAGTTCCAGACCGTTGA 588


RESULT 8
BY738990
LOCUS       BY738990                653 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY738990 RIKEN full-length enriched, 17 days pregnant adult female
            amnion Mus musculus cDNA clone I920028B19 5', mRNA sequence.
ACCESSION   BY738990
VERSION     BY738990.1  GI:27162492
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 653)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .653
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="I920028B19"
                      /sex="female"
                      /tissue_type="amnion"
                      /dev_stage="17 days pregnant adult"
                      /clone_lib="RIKEN full-length enriched, 17 days pregnant
                      adult female amnion"
BASE COUNT      149 a    199 c    180 g    122 t      3 others
ORIGIN

  Query Match           78.0%;  Score 183.2;  DB 14;  Length 653;
  Best Local Similarity   89.4%;  Pred. No. 1e-42;
  Matches  210;  Conservative    0;  Mismatches   16;  Indels    9;  Gaps    1;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        403 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCANAGGAACAGAGTG 462

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              |  ||||||||||||||||          |||||||||||||| ||||||||||||||||
Db        463 GGCACAGTACCAATGGCAGC---------GAGTCTTCCAAGAACCGCACAGTCTCTGGTG 513

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||||||||||||||||||  | ||||||||||||||||  |||||||||||| || |
Db        514 GGCAGTATGTTGTGGCTGCTACCCCCAACTTACAGAACCAGNCAGTTCTGACAGGTCTCC 573

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| ||||||||||||||||||||||||||||||| |||||||  ||||
Db        574 CTGGAGTAATGCCTAATATTCAGTATCAAGTAATCCCACAGNTCCAGACTATTGA 628


RESULT 9
BU702181
LOCUS       BU702181                586 bp    mRNA    linear   EST 09-OCT-2002
DEFINITION  UI-M-FI0-byr-a-11-0-UI.r1 NIH_BMAP_FI0 Mus musculus cDNA clone
            IMAGE: 5698066 5', mRNA sequence.
ACCESSION   BU702181
VERSION     BU702181.1  GI:23626729
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1  (bases 1 to 586)
  AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Dr. Jim Lin, University of Iowa
           cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
           cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
           DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
           Clone Distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
           This clone was contributed by the Brain Molecular Anatomy Project
          (BMAP)
          Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 5698066"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FI0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is CAGCCACGAC. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      143 a    167 c    161 g    115 t
ORIGIN

  Query Match             77.9%;  Score 183;  DB 13;  Length 586;
  Best Local Similarity   89.8%;  Pred. No. 1.1e-42;
  Matches  211;  Conservative    0;  Mismatches   15;  Indels    9;  Gaps    1;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        196 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAGGGAACAGAGTG 255

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              | | |||||||||||| |||          |||||||||||||| ||||||||||||||||
Db        256 GGAACAGTACCAATGGGAGC---------GAGTCTTCCAAGAACCGCACAGTCTCTGGTG 306

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||| ||||||||||||  | |||| ||||||||||||||||||||||||||| || |
Db        307 GGCAGTGTGTTGTGGCTGCTACCCCCACCTTACAGAACCAGCAAGTTCTGACAGGTCTCC 366

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||| ||||||| ||||||||||||||||||||||||||||||||| |||||
Db        367 CTGGAGTAATGCCTACTATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 421


RESULT 10
AW445976
LOCUS       AW445976               416 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  83265 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW445976
VERSION     AW445976.1  GI:6987761
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 45  row: B  column: 16
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 1BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      113 a    114 c    104 g     85 t
ORIGIN

  Query Match             67.0%;  Score 157.4;  DB 9;  Length 416;
  Best Local Similarity   93.7%;  Pred. No. 2.8e-35;
  Matches  164;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              |||||||||||||| |||  ||||||||||||||||||||||||||||| ||||| ||||
Db          1 GCAGCAGTACCAACGGCGACAATGGCAGTGAGTCTTCCAAGAATCGCACGGTCTCCGGTG 60

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              |||||||||||||||||| || ||||| | ||||||||||||| ||||| |||||||
Db         61 GGCAGTATGTTGTGGCTGCCACTTCCAACCTTCAGAACCAGCAAGTCCTGACCGGACTAC 120

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 175


RESULT 11
CB811553
LOCUS       CB811553               404 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  AMGNNUC:SRPB2-00245-E5-A srpb2 (10220) Rattus norvegicus cDNA clone
            srpb2-00245-e5 5', mRNA sequence.
ACCESSION   CB811553
VERSION     CB811553.1  GI:29934519
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00245  row: e  column: 5.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="srpb2-00245-e5"
                     /tissue_type="prostate tissue"
                     /clone_lib="srpb2 (10220)"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; rat
                     prostate normalized double selected poly(A+) mRNA size
                     fraction > 1 kb"
BASE COUNT       94 a    113 c     93 g    102 t      2 others
ORIGIN

  Query Match             62.6%;  Score 147.2;  DB 14;  Length 404;
  Best Local Similarity   94.4%;  Pred. No. 2.6e-32;
  Matches  152;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy         75 GGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTG 134
              ||||||||||||||||||||||||  || ||||||||||||||||||||||||||||||
Db          1 GGCAGCAATGGCAGTGAGTCTTCCANTAACCGCACAGTCTCTGGTGGGCAGTATGTTGTG 60

Qy        135 GCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTACCTGGAGTGATGCCT 194
              |||||  | ||||||||||||||||||||||||||||||||| |||||||||| ||||||
Db         61 GCTGCTACCCCCAACTTACAGAACCAGCAAGTTCTGACAGGTCTACCTGGAGTAATGCCT 120

Qy        195 AATATTCAGTATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              |||||||||||||||||||||||||||||||||| |||||
Db        121 AATATTCAGTATCAAGTAATCCCACAGTTCCAGACTGTTGA 161


RESULT 12
BM427106
LOCUS       BM427106                 681 bp    mRNA    linear   EST 30-JAN-2002
DEFINITION  pgf2n.pk005.l17 Normalized Chicken Abdominal Fat Library (pgf2n)
            Gallus gallus cDNA clone pgf2n.pk005.l17 5' similar to
            dbj|BAB13476.1 (AB039286) transcription factor Sp1 [Homo sapiens],
            mRNA sequence.
ACCESSION   BM427106
VERSION     BM427106.1  GI:18432006
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Cogburn,L.A., Morgan,R. and Burnside,J.
  TITLE     ESTs from Normalized Chicken fat cDNA library-USDA/IFAFS Animal
            Genome Project
  JOURNAL   Unpublished
COMMENT     Contact: Larry A. Cogburn
            University of Delaware
            Townsend Hall, Newark, DE 19717, USA
            Tel: 302-831-1335
            Fax: 302-831-2822
            Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Commercial broiler, Ottawa Research Centre,
                     leghorn"
                     /db_xref="taxon:9031"
                     /clone="pgf2n.pk005.l17"
                     /sex="Male and Female"
                     /tissue_type="Abdominal Fat"
                     /dev_stage="Embryonic (d18,d19); post-hatch (d1,w3,w7,w9
                     ,w16,1yr)"
                     /lab_host="E. coli EMDH10B"
                     /clone_lib="Normalized Chicken Abdominal Fat Library
                     (pgf2n)"
                     /note="Vector: pCMVSPORT6; Library made from equivalent
                     pools of total RNA isolated  from each developmental age
                     (across strains); Single pass sequencing from 5'-end"
BASE COUNT       146 a    183 c    265 g     82 t      5 others
ORIGIN

  Query Match             32.1%;  Score 75.4;  DB 12;  Length 681;
  Best Local Similarity   61.4%;  Pred. No. 3e-11;
  Matches  140;  Conservative    0;  Mismatches   76;  Indels   12;  Gaps    1;

Qy          1 ATGGCTGGCAGATCATCTCTTCCTCCTCTGGGGCTACCCCTACCTCAAAGGAACAGAGTG 60
              | |||||||||||||||||  ||  |||     |  |  |  ||||  |||||  ||| |
Db        466 ACGGCTGGCAGATCATCTCCACCGGCTCCACCACGCCGACGCCCTCCAAGGAGCAGGGCA 525

Qy         61 GCAGCAGTACCAATGGCAGCAATGGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTG 120
              |||   | | |||              | | ||  | |||||  |  ||  | ||       |
Db        526 GCAATGGCAGCAA------------CGGAGATCCCTCCAAAAGCCGTCCCGTGAGCACGG 573

Qy        121 GGCAGTATGTTGTGGCTGCCGCTCCCAACTTACAGAACCAGCAAGTTCTGACAGGACTAC 180
              ||| ||| || ||  | || ||| ||||| | ||||||||  |      || || || || |
Db        574 GGCCGTACGTGGTCACGGCTGCTTCCAACCTGCAGAACCANCNNNNGCTCACGGGCCTGC 633

Qy        181 CTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCCACAGTTCCAGA 228
              | || || || || || || ||||| || || || || ||||||||||
Db        634 CGGGCGTCATCCCCAACATCCAGTACCAGGTCATTCCCCAGTTCCAGA 681


RESULT 13
BG997913/c
LOCUS       BG997913                 504 bp    mRNA    linear   EST 13-JUN-2001
DEFINITION  PM1-HT1298-090401-002-a03 HT1298 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG997913
VERSION     BG997913.1  GI:14401983
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM1&t2=PM1-HT1298-
            090401-002-a03&t3=2001-04-09&t4=1)
            Seq primer: puc 18 forward
```

```
           High quality sequence start: 8
           High quality sequence stop: 504.
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT1298"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      113 a    115 c    148 g    127 t      1 others
ORIGIN

  Query Match             29.8%;  Score 70;  DB 12;  Length 504;
  Best Local Similarity   93.6%;  Pred. No. 9.6e-10;
  Matches   73;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        158 CCAGCAAGTTCTGACAGGACTACCTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCC 217
              ||    ||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        494 CCCAGAACCTCTGACAGGACTACCTGGAGTGATGCCTAATATTCAGTATCAAGTAATCCC 435

Qy        218 ACAGTTCCAGACCGTTGA 235
              ||||||||||||||||||
Db        434 ACAGTTCCAGACCGTTGA 417


RESULT 14
BU906803
LOCUS       BU906803                880 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  AGENCOURT_10456317 NICHD_XGC_Emb1 Xenopus laevis cDNA clone
            IMAGE:6631848 5', mRNA sequence.
ACCESSION   BU906803
VERSION     BU906803.1  GI:24088717
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 880)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14200  row: d  column: 24
            High quality sequence stop: 647.
FEATURES             Location/Qualifiers
     source          1. .880
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6631848"
                     /tissue_type="embryo (stage 10)"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_Emb1"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 1.55 kb. Constructed by Life Technologies. Note: This
                     is a Xenopus Gene Collection (XGC) library."
BASE COUNT      238 a    243 c    190 g    207 t      2 others
ORIGIN

  Query Match             28.6%;  Score 67.2;  DB 13;  Length 880;
  Best Local Similarity   69.7%;  Pred. No. 8.4e-09;
  Matches  106;  Conservative    0;  Mismatches   43;  Indels    3;  Gaps    1;

Qy         84 GGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTGCCGCT 143
              ||   ||  ||||| || || ||| || |  |   ||||||||   ||||||  || ||
Db        300 GGAGATGCTTCTTCTAAAAACCGCCCAATAGCCCCTGGGCAGT---TTGTGGTAGCTGCA 356

Qy        144 CCCAACTTACAGAACCAGCAAGTTCTGACAGGACTACCTGGAGTGATGCCTAATATTCAG 203
              ||||   | || || ||||| ||| ||||  | | |   | ||||||| || ||||||
Db        357 CCCAGTGTGCAAAATCAGCAGGTTTTGACCAGTTTGCAGAGTGTGATGCCCAACATTCAG 416

Qy        204 TATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              || ||||| || ||||| |||||||| |||||
Db        417 TACCAAGTCATACCACAATTCCAGACTGTTGA 448


RESULT 15
CA787807
LOCUS       CA787807                928 bp    mRNA    linear   EST 04-DEC-2002
DEFINITION  AGENCOURT_11031172 NICHD_XGC_001 Xenopus laevis cDNA clone
            IMAGE:6859147 5', mRNA sequence.
ACCESSION   CA787807
VERSION     CA787807.1  GI:26031222
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14471  row: g  column: 18
            High quality sequence stop: 744.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6859147"
                     /tissue_type="oocytes"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_001"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.2 kb. Constructed by Life Technologies."
BASE COUNT      251 a    249 c    201 g    223 t      4 others
ORIGIN

  Query Match             28.6%;  Score 67.2;  DB 14;  Length 928;
  Best Local Similarity   69.7%;  Pred. No. 8.7e-09;
  Matches  106;  Conservative    0;  Mismatches   43;  Indels    3;  Gaps    1;

Qy         84 GGCAGTGAGTCTTCCAAGAATCGCACAGTCTCTGGTGGGCAGTATGTTGTGGCTGCCGCT 143
              ||   ||  ||||| || || ||| || |  |   ||||||||   ||||||  || ||
Db        329 GGAGATGCTTCTTCTAAAAACCGCCCAATAGCCCCTGGGCAGT---TTGTGGTAGCTGCA 385
```

```
Qy        144 CCCAACTTACAGAACCAGCAAGTTCTGACAGGACTACCTGGAGTGATGCCTAATATTCAG 203
              ||||    | || || ||||| ||| ||||  |  | |    | |||||||| || ||||||
Db        386 CCCAGTGTGCAAAATCAGCAGGTTTTGACCAGTTTGCAGAGTGTGATGCCCAACATTCAG 445

Qy        204 TATCAAGTAATCCCACAGTTCCAGACCGTTGA 235
              || ||||| || ||||| |||||||| |||||
Db        446 TACCAAGTCATACCACAATTCCAGACTGTTGA 477
```

Search completed: February 18, 2004, 16:09:26
Job time : 2374.2 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 15:26:16 ; Search time 1542 Seconds
(without alignments)
583.665 Million cell updates/sec

Title: US-10-026-341A-2
Perfect score: 22
Sequence: 1 attcgatcggggcggggcgagc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	22	100.0	22	6	AR060640	AR060640 Sequence
2	22	100.0	22	6	AR070792	AR070792 Sequence
3	22	100.0	22	6	AX195274	AX195274 Sequence
4	22	100.0	22	6	AX526281	AX526281 Sequence
5	22	100.0	22	6	BD105476	BD105476 Prolifera
6	22	100.0	48	6	AX377574	AX377574 Sequence
c 7	22	100.0	48	6	AX377574	AX377574 Sequence
8	22	100.0	49	6	AX127457	AX127457 Sequence
c 9	22	100.0	49	6	AX127457	AX127457 Sequence
10	22	100.0	49	6	AX127459	AX127459 Sequence
c 11	22	100.0	49	6	AX127459	AX127459 Sequence
12	22	100.0	49	6	AX377573	AX377573 Sequence
13	21	95.5	21	6	AX104633	AX104633 Sequence
c 14	21	95.5	21	6	AX104634	AX104634 Sequence
15	21	95.5	21	6	AX355087	AX355087 Sequence
c 16	21	95.5	21	6	AX355201	AX355201 Sequence
17	21	95.5	21	6	AX547686	AX547686 Sequence
c 18	21	95.5	21	6	AX547687	AX547687 Sequence
19	20	90.9	20	6	A89791	A89791 Sequence 13
20	20	90.9	20	6	A90878	A90878 Sequence 13
21	20	90.9	20	6	AX023404	AX023404 Sequence
22	20	90.9	20	6	AX455627	AX455627 Sequence
23	20	90.9	20	6	BD056782	BD056782 Pharmaceu
24	19	86.4	46	6	I72381	I72381 Sequence 12
25	19	86.4	46	6	I72382	I72382 Sequence 13
c 26	18.8	85.5	49	6	AX377573	AX377573 Sequence
27	18.4	83.6	22	6	E07877	E07877 Synthetic n
c 28	18	81.8	46	6	I72381	I72381 Sequence 12
c 29	18	81.8	46	6	I72382	I72382 Sequence 13
c 30	17.8	80.9	181162	5	AC144487	AC144487 Gasterost
31	17.8	80.9	274015	2	AC127795	AC127795 Rattus no
32	17.8	80.9	301660	2	AC112802	AC112802 Rattus no
33	17.4	79.1	1119	10	BC016449	BC016449 Mus muscu
34	17.4	79.1	1123	9	HSA243936	AJ243936 Homo sapi
35	17.4	79.1	7125	1	AF335479	AF335479 Agrobacte
36	17.4	79.1	9646	6	AX346590	AX346590 Sequence
c 37	17.4	79.1	30377	9	HSLUCA1	Z74618 Human DNA s
c 38	17.4	79.1	106278	2	RN520N23	AL732653 Rattus no
39	17.4	79.1	126323	8	AC132215	AC132215 Genomic s
40	17.4	79.1	144969	8	AC137547	AC137547 Oryza sat
c 41	17.4	79.1	158266	2	AC092556	AC092556 Oryza sat
c 42	17.4	79.1	184461	2	AC118476	AC118476 Mus muscu
c 43	17.4	79.1	192324	9	AC096920	AC096920 Homo sapi
c 44	17.4	79.1	201964	10	MMMHC29N7	AF030001 Mus muscu
c 45	17.4	79.1	201986	10	AC006289	AC006289 Mus muscu

ALIGNMENTS

RESULT 1
AR060640
LOCUS AR060640 22 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 12 from patent US 5840832.
ACCESSION AR060640
VERSION AR060640.1 GI:5987090
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 22)
AUTHORS Ono,S.Jeremy. and Strominger,J.L.
TITLE Transcription factor regulating MHC expression, CDNA and genomic clones encoding same and retroviral expression constructs thereof
JOURNAL Patent: US 5840832-A 12 24-NOV-1998;

```
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="unknown"
BASE COUNT        3 a      5 c     11 g      3 t
ORIGIN

  Query Match             100.0%;  Score 22;  DB 6;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 2
AR070792
LOCUS       AR070792                 22 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 12 from patent US 5908762.
ACCESSION   AR070792
VERSION     AR070792.1  GI:7221680
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   Ono,S.Jeremy. and Strominger,J.L.
  TITLE     Transcription factor regulating MHC expression CDNA and genomic
            clones encoding same and retroviral expression constructs thereof
  JOURNAL   Patent: US 5908762-A 12 01-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="unknown"
BASE COUNT        3 a      5 c     11 g      3 t
ORIGIN

  Query Match             100.0%;  Score 22;  DB 6;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 3
AX195274
LOCUS       AX195274                 22 bp    DNA     linear   PAT 28-AUG-2001
DEFINITION  Sequence 10 from Patent WO0151671.
ACCESSION   AX195274
VERSION     AX195274.1  GI:15385825
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Mccarthy,J. and Cordell,B.
  TITLE     Methods for identifying inhibitors of neuronal degeneration
  JOURNAL   Patent: WO 0151671-A 10 19-JUL-2001;
            Scios Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic Oligonucleotide"
BASE COUNT        3 a      5 c     11 g      3 t
ORIGIN

  Query Match             100.0%;  Score 22;  DB 6;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 4
AX526281
LOCUS       AX526281                 22 bp    DNA     linear   PAT 21-NOV-2002
DEFINITION  Sequence 2 from Patent WO02066071.
ACCESSION   AX526281
VERSION     AX526281.1  GI:25171091
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Mauviel,A.
  TITLE     Blocking sp1 transcription factor broadly inhibits extracellular
            matrix gene expression in vitro and in vivo: implications for the
            treatments of tissue fibrosis
  JOURNAL   Patent: WO 02066071-A 2 29-AUG-2002;
            Thomas Jefferson University (US)
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
                     /note="oligonucleotide"
BASE COUNT        3 a      5 c     11 g      3 t
ORIGIN

  Query Match             100.0%;  Score 22;  DB 6;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 5
BD105476
LOCUS       BD105476                 22 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Proliferating agents and apoptosis-suppressing agents for
            pancreatic beta-cells of the islets of langerhans, and screening of
            candidate compounds thereof.
ACCESSION   BD105476
VERSION     BD105476.1  GI:22651050
KEYWORDS    WO 0193899-A/9.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   Okamoto,H.
  TITLE     Proliferating agents and apoptosis-suppressing agents for
            pancreatic beta-cells of the islets of langerhans, and screening of
            candidate compounds thereof
  JOURNAL   Patent: WO 0193899-A 9 13-DEC-2001;
            HIROSHI OKAMOTO
COMMENT     OS   Artificial Sequence
            PN   WO 0193899-A/9
            PD   13-DEC-2001
            PF   01-JUN-2001 WO 2001JP004660
            PR   02-JUN-2000 JP 00P   170447,28-FEB-2001 JP 01P    54072 PI
            HIROSHI OKAMOTO
            PC   A61K38/20,A61K31/573,A61K38/22,A61K48/00,A61K45/00,A61K31/455,
            PC   A61K31/165,
            PC   A61P43/00,A61P3/10,G01N33/15,G01N33/50,C12Q1/02//C12N15/12, PC
              C12N5/06
            CC   Description of Artificial Sequence:an artificially synthesized
            CC   oligonucleotide sequence
            FH   Key             Location/Qualifiers
            FT   source          1. .22
```

```
          FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
BASE COUNT        3 a      5 c     11 g      3 t
ORIGIN

  Query Match             100.0%;  Score 22;  DB 6;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 6
AX377574
LOCUS       AX377574                  48 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 51 from Patent WO0212553.
ACCESSION   AX377574
VERSION     AX377574.1  GI:19573760
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Kappel,A., Polakowski,T., Pignot,M., Windhab,N., Behrensdorf,H. and
            Muth,J.
  TITLE     Method for detecting mutations in nucleotide sequences
  JOURNAL   Patent: WO 0212553-A 51 14-FEB-2002;
            Nanogen Recognomics GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Beschreibung der kunstlichen Sequenz:
                     Hairpin-Oligonucleotid"
BASE COUNT        6 a     16 c     16 g     10 t
ORIGIN

  Query Match             100.0%;  Score 22;  DB 6;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 7
AX377574/c
LOCUS       AX377574                  48 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 51 from Patent WO0212553.
ACCESSION   AX377574
VERSION     AX377574.1  GI:19573760
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Kappel,A., Polakowski,T., Pignot,M., Windhab,N., Behrensdorf,H. and
            Muth,J.
  TITLE     Method for detecting mutations in nucleotide sequences
  JOURNAL   Patent: WO 0212553-A 51 14-FEB-2002;
            Nanogen Recognomics GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Beschreibung der kunstlichen Sequenz:
                     Hairpin-Oligonucleotid"
BASE COUNT        6 a     16 c     16 g     10 t
ORIGIN

  Query Match             100.0%;  Score 22;  DB 6;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db         48 ATTCGATCGGGGCGGGGCGAGC 27


RESULT 8
AX127457
LOCUS       AX127457                  49 bp    DNA     linear   PAT 15-MAY-2001
DEFINITION  Sequence 1 from Patent WO0131057.
ACCESSION   AX127457
VERSION     AX127457.1  GI:14134020
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Muth,J. and Windhab,N.
  TITLE     Double-strand nucleic acid probes and the use thereof
  JOURNAL   Patent: WO 0131057-A 1 03-MAY-2001;
            Aventis Research & Technologies GmbH & Co KG (DE)
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Detektorsequenz mit einem Amino-Modifier
                     dT-Baustein an Position 25"
BASE COUNT        6 a     16 c     16 g     11 t
ORIGIN

  Query Match             100.0%;  Score 22;  DB 6;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 9
AX127457/c
LOCUS       AX127457                  49 bp    DNA     linear   PAT 15-MAY-2001
DEFINITION  Sequence 1 from Patent WO0131057.
ACCESSION   AX127457
VERSION     AX127457.1  GI:14134020
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Muth,J. and Windhab,N.
  TITLE     Double-strand nucleic acid probes and the use thereof
  JOURNAL   Patent: WO 0131057-A 1 03-MAY-2001;
            Aventis Research & Technologies GmbH & Co KG (DE)
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Detektorsequenz mit einem Amino-Modifier
                     dT-Baustein an Position 25"
BASE COUNT        6 a     16 c     16 g     11 t
ORIGIN
```

```
  Query Match            100.0%;  Score 22;  DB 6;  Length 49;
  Best Local Similarity  100.0%;  Pred. No. 45;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db         49 ATTCGATCGGGGCGGGGCGAGC 28


RESULT 10
AX127459
LOCUS       AX127459                  49 bp    DNA     linear   PAT 15-MAY-2001
DEFINITION  Sequence 3 from Patent WO0131057.
ACCESSION   AX127459
VERSION     AX127459.1  GI:14134028
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Muth,J. and Windhab,N.
  TITLE     Double-strand nucleic acid probes and the use thereof
  JOURNAL   Patent: WO 0131057-A 3 03-MAY-2001;
            Aventis Research & Technologies GmbH & Co KG (DE)
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Detektorsequenz mit einem Amino-Modifier
                     dT-Baustein an Position 27"
BASE COUNT        6 a     16 c     16 g     11 t
ORIGIN

  Query Match            100.0%;  Score 22;  DB 6;  Length 49;
  Best Local Similarity  100.0%;  Pred. No. 45;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 11
AX127459/c
LOCUS       AX127459                  49 bp    DNA     linear   PAT 15-MAY-2001
DEFINITION  Sequence 3 from Patent WO0131057.
ACCESSION   AX127459
VERSION     AX127459.1  GI:14134028
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Muth,J. and Windhab,N.
  TITLE     Double-strand nucleic acid probes and the use thereof
  JOURNAL   Patent: WO 0131057-A 3 03-MAY-2001;
            Aventis Research & Technologies GmbH & Co KG (DE)
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Detektorsequenz mit einem Amino-Modifier
                     dT-Baustein an Position 27"
BASE COUNT        6 a     16 c     16 g     11 t
ORIGIN

  Query Match            100.0%;  Score 22;  DB 6;  Length 49;
  Best Local Similarity  100.0%;  Pred. No. 45;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db         49 ATTCGATCGGGGCGGGGCGAGC 28


RESULT 12
AX377573
LOCUS       AX377573                  49 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 50 from Patent WO0212553.
ACCESSION   AX377573
VERSION     AX377573.1  GI:19573759
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Kappel,A., Polakowski,T., Pignot,M., Windhab,N., Behrensdorf,H. and
            Muth,J.
  TITLE     Method for detecting mutations in nucleotide sequences
  JOURNAL   Patent: WO 0212553-A 50 14-FEB-2002;
            Nanogen Recognomics GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Beschreibung der kunstlichen Sequenz:
                     Hairpin-Oligonucleotid"
BASE COUNT        6 a     14 c     16 g     13 t
ORIGIN

  Query Match            100.0%;  Score 22;  DB 6;  Length 49;
  Best Local Similarity  100.0%;  Pred. No. 45;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 13
AX104633
LOCUS       AX104633                  21 bp    DNA     linear   PAT 30-APR-2001
DEFINITION  Sequence 825 from Patent WO0122972.
ACCESSION   AX104633
VERSION     AX104633.1  GI:13920830
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Krieg,A.M., Schetter,C. and Vollmer,J.C.
  TITLE     Immunostimulatory nucleic acids
  JOURNAL   Patent: WO 0122972-A 825 05-APR-2001;
            UNIVERSITY OF IOWA RESEARCH FOUNDATION (US) ; Coley Pharmaceutical
            GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
BASE COUNT        3 a      4 c     11 g      3 t
ORIGIN

  Query Match            95.5%;  Score 21;  DB 6;  Length 21;
  Best Local Similarity  100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAG 21
```

```
RESULT 14
AX104634/c
LOCUS       AX104634                 21 bp    DNA     linear   PAT 30-APR-2001
DEFINITION  Sequence 826 from Patent WO0122972.
ACCESSION   AX104634
VERSION     AX104634.1  GI:13920831
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Krieg,A.M.; Schetter,C. and Vollmer,J.C.
  TITLE     Immunostimulatory nucleic acids
  JOURNAL   Patent: WO 0122972-A 826 05-APR-2001;
            UNIVERSITY OF IOWA RESEARCH FOUNDATION (US) ; Coley Pharmaceutical
            GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
BASE COUNT        3 a      11 c       4 g       3 t
ORIGIN

  Query Match             95.5%;  Score 21;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db         21 ATTCGATCGGGGCGGGGCGAG 1


RESULT 15
AX355087
LOCUS       AX355087                 21 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 115 from Patent WO0197843.
ACCESSION   AX355087
VERSION     AX355087.1  GI:18619754
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Weiner,G. and Hartmann,G.
  TITLE     Methods for enhancing antibody-induced cell lysis and treating
            cancer
  JOURNAL   Patent: WO 0197843-A 115 27-DEC-2001;
            UNIVERSITY OF IOWA RESEARCH FOUNDATION (US)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide-phosphodiester backbone"
BASE COUNT        3 a       4 c      11 g       3 t
ORIGIN

  Query Match             95.5%;  Score 21;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAG 21


RESULT 16
AX355201/c
LOCUS       AX355201                 21 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 229 from Patent WO0197843.
ACCESSION   AX355201
VERSION     AX355201.1  GI:18619868
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Weiner,G. and Hartmann,G.
  TITLE     Methods for enhancing antibody-induced cell lysis and treating
            cancer
  JOURNAL   Patent: WO 0197843-A 229 27-DEC-2001;
            UNIVERSITY OF IOWA RESEARCH FOUNDATION (US)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide-phosphodiester backbone"
BASE COUNT        3 a      11 c       4 g       3 t
ORIGIN

  Query Match             95.5%;  Score 21;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db         21 ATTCGATCGGGGCGGGGCGAG 1


RESULT 17
AX547686
LOCUS       AX547686                 21 bp    DNA     linear   PAT 26-NOV-2002
DEFINITION  Sequence 825 from Patent WO02053141.
ACCESSION   AX547686
VERSION     AX547686.1  GI:25812830
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Bratzler,R.L.
  TITLE     Inhibition of angiogenesis by nucleic acids
  JOURNAL   Patent: WO 02053141-A 825 11-JUL-2002;
            Coley Pharmaceutical Group, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic Sequence"
BASE COUNT        3 a       4 c      11 g       3 t
ORIGIN

  Query Match             95.5%;  Score 21;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAG 21


RESULT 18
AX547687/c
LOCUS       AX547687                 21 bp    DNA     linear   PAT 26-NOV-2002
DEFINITION  Sequence 826 from Patent WO02053141.
ACCESSION   AX547687
VERSION     AX547687.1  GI:25812831
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Bratzler,R.L.
```

```
  TITLE     Inhibition of angiogenesis by nucleic acids
  JOURNAL   Patent: WO 02053141-A 826 11-JUL-2002;
            Coley Pharmaceutical Group, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic Sequence"
BASE COUNT        3 a      11 c      4 g      3 t
ORIGIN

  Query Match             95.5%;  Score 21;  DB 6;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db         21 ATTCGATCGGGGCGGGGCGAG 1


RESULT 19
A89791
LOCUS       A89791                    20 bp    DNA     linear   PAT 22-JAN-2000
DEFINITION  Sequence 13 from Patent WO9832462.
ACCESSION   A89791
VERSION     A89791.1  GI:6738305
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Lipford,G.B. and Heeg,K.
  TITLE     PHARMACEUTICAL COMPOSITIONS COMPRISING A POLYNUCLEOTIDE AND
            OPTIONALLY AN ANTIGEN ESPECIALLY FOR VACCINATION
  JOURNAL   Patent: WO 9832462-A 13 30-JUL-1998;
            LIPFORD GRAYSON B (DE); HEEG KLAUS (DE)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT        2 a      5 c      11 g      2 t
ORIGIN

  Query Match             90.9%;  Score 20;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||
Db          1 TCGATCGGGGCGGGGCGAGC 20


RESULT 20
A90878
LOCUS       A90878                    20 bp    DNA     linear   PAT 22-JAN-2000
DEFINITION  Sequence 13 from Patent EP0855184.
ACCESSION   A90878
VERSION     A90878.1  GI:6739281
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Heeg,K.P. and Lipford,G.B.
  TITLE     Pharmaceutical composition comprising a polynucleotide and an
            antigen especially for vaccination
  JOURNAL   Patent: EP 0855184-A 13 29-JUL-1998;
            HEEG KLAUS PROF DR (DE); LIPFORD GRAYSON B DR (DE)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT        2 a      5 c      11 g      2 t
ORIGIN

  Query Match             90.9%;  Score 20;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||
Db          1 TCGATCGGGGCGGGGCGAGC 20


RESULT 21
AX023404
LOCUS       AX023404                  20 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 19 from Patent WO0014217.
ACCESSION   AX023404
VERSION     AX023404.1  GI:10183804
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Lipford,G.B., Heeg,K. and Wagner,H.
  TITLE     G-motif oligonucleotides and uses thereof
  JOURNAL   Patent: WO 0014217-A 19 16-MAR-2000;
            LIPFORD GRAYSON B (DE) ; HEEG KLAUS (DE) ; WAGNER HERMANN (DE) ;
            CPG IMMUNOPHARMACEUTICALS GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="synthetic, no natural origin"
BASE COUNT        2 a      5 c      11 g      2 t
ORIGIN

  Query Match             90.9%;  Score 20;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||
Db          1 TCGATCGGGGCGGGGCGAGC 20


RESULT 22
AX455627
LOCUS       AX455627                  20 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 104 from Patent WO0222809.
ACCESSION   AX455627
VERSION     AX455627.1  GI:21714695
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Bauer,S., Lipford,G. and Wagner,H.
  TITLE     Process for high throughput screening of cpg-based
            immuno-agonist/antagonist
  JOURNAL   Patent: WO 0222809-A 104 21-MAR-2002;
            Coley Pharmaceutical GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide"
BASE COUNT        2 a      5 c      11 g      2 t
ORIGIN
```

```
  Query Match             90.9%;  Score 20;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||
Db          1 TCGATCGGGGCGGGGCGAGC 20


RESULT 23
BD056782
LOCUS       BD056782                  20 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Pharmaceutical composition comprising a polynucleotide and an
            antigen especially for vaccination.
ACCESSION   BD056782
VERSION     BD056782.1  GI:22602388
KEYWORDS    JP 2001508780-A/12.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Wagner,H., Lipford,G.B. and Heeg,K.
  TITLE     Pharmaceutical composition comprising a polynucleotide and an
            antigen especially for vaccination
  JOURNAL   Patent: JP 2001508780-A 12 03-JUL-2001;
            HERMANN WAGNER,GRAYSON B LIPFORD,KLAUS HEEG
COMMENT     PN   JP 2001508780-A/12
            PD   03-JUL-2001
            PF   23-JAN-1998 JP 1998531592
            PR   23-JAN-1997 EP   97101019.4
            PI   HERMANN WAGNER,GRAYSON B LIPFORD,KLAUS HEEG
            PC   A61K39/39,A61K31/7088,A61K39/00,A61P37/04,C12N15/09,C12N15/00
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES                Location/Qualifiers
     source             1. .20
                        /organism="synthetic construct"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:32630"
BASE COUNT        2 a      5 c     11 g      2 t
ORIGIN

  Query Match             90.9%;  Score 20;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||
Db          1 TCGATCGGGGCGGGGCGAGC 20


RESULT 24
I72381
LOCUS       I72381                    46 bp    DNA     linear   PAT 03-APR-1998
DEFINITION  Sequence 12 from patent US 5683985.
ACCESSION   I72381
VERSION     I72381.1  GI:3008520
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Chu,B.Chen.Fei. and Orgel,L.
  TITLE     Oligonucleotide decoys and methods relating thereto
  JOURNAL   Patent: US 5683985-A 12 04-NOV-1997;
FEATURES                Location/Qualifiers
     source             1. .46
                        /organism="unknown"
BASE COUNT        3 a      21 c     15 g      7 t
ORIGIN

  Query Match             86.4%;  Score 19;  DB 6;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||
Db         28 TCGATCGGGGCGGGGCGAG 46


RESULT 25
I72382
LOCUS       I72382                    46 bp    DNA     linear   PAT 03-APR-1998
DEFINITION  Sequence 13 from patent US 5683985.
ACCESSION   I72382
VERSION     I72382.1  GI:3008521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Chu,B.Chen.Fei. and Orgel,L.
  TITLE     Oligonucleotide decoys and methods relating thereto
  JOURNAL   Patent: US 5683985-A 13 04-NOV-1997;
FEATURES                Location/Qualifiers
     source             1. .46
                        /organism="unknown"
BASE COUNT        3 a      21 c     15 g      7 t
ORIGIN

  Query Match             86.4%;  Score 19;  DB 6;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||
Db         28 TCGATCGGGGCGGGGCGAG 46


RESULT 26
AX377573/c
LOCUS       AX377573                  49 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 50 from Patent WO0212553.
ACCESSION   AX377573
VERSION     AX377573.1  GI:19573759
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Kappel,A., Polakowski,T., Pignot,M., Windhab,N., Behrensdorf,H. and
            Muth,J.
  TITLE     Method for detecting mutations in nucleotide sequences
  JOURNAL   Patent: WO 0212553-A 50 14-FEB-2002;
            Nanogen Recognomics GmbH (DE)
FEATURES                Location/Qualifiers
     source             1. .49
                        /organism="synthetic construct"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:32630"
                        /note="Beschreibung der kunstlichen Sequenz:
                        Hairpin-Oligonucleotid"
BASE COUNT        6 a      14 c     16 g      13 t
ORIGIN

  Query Match             85.5%;  Score 18.8;  DB 6;  Length 49;
  Best Local Similarity   90.9%;  Pred. No. 1e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              |||||||||||||  |||||||
Db         49 ATTCGATCGGGGCAAGGCGAGC 28


RESULT 27
```

```
E07877
LOCUS       E07877                    22 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  Synthetic nucleotides for DNA binding protein analysis.
ACCESSION   E07877
VERSION     E07877.1  GI:2176010
KEYWORDS    JP 1994201692-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 22)
  AUTHORS   Uno,I., Isohara,T., Igarashi,K., Miki,N. and Ou,S.
  TITLE     PROBE FOR DETECTING OR MEASURING DNA-BINDING PROTEIN AND METHOD FOR
            DETECTING OR MEASURING DNA-BINDING PROTEIN USING IT
  JOURNAL   Patent: JP 1994201692-A 2 22-JUL-1994;
            NIPPON STEEL CORP, NIPPON STEEL CHEM CO LTD
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1994201692-A/2
            PD   22-JUL-1994
            PF   22-SEP-1993 JP 1993257482
            PR   22-SEP-1992 JP 92P   278126
            PI   UNO ISAO, ISOHARA TOSHIO, IGARASHI KATSUHIDE, MIKI NAOMASA, PI
              OU SHIYOUHEI
            PC   G01N33/53//C12Q1/68;
            CC   strandedness: Single;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .22
            FT                   /organism='Artificial sequences'.
FEATURES             Location/Qualifiers
     source          1. .22
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT        3 a      5 c     11 g      3 t
ORIGIN

  Query Match           83.6%;  Score 18.4;  DB 6;  Length 22;
  Best Local Similarity   95.0%;  Pred. No. 1.6e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||| |
Db          3 TCGATCGGGGCGGGGCGATC 22


RESULT 28
I72381/c
LOCUS       I72381                    46 bp    DNA     linear   PAT 03-APR-1998
DEFINITION  Sequence 12 from patent US 5683985.
ACCESSION   I72381
VERSION     I72381.1  GI:3008520
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Chu,B.Chen.Fei. and Orgel,L.
  TITLE     Oligonucleotide decoys and methods relating thereto
  JOURNAL   Patent: US 5683985-A 12 04-NOV-1997;
FEATURES             Location/Qualifiers
     source          1. .46
                     /organism="unknown"
BASE COUNT        3 a      21 c     15 g      7 t
ORIGIN

  Query Match           81.8%;  Score 18;  DB 6;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+03;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||
Db         24 GATCGGGGCGGGGCGAGC 7


RESULT 29
I72382/c
LOCUS       I72382                    46 bp    DNA     linear   PAT 03-APR-1998
DEFINITION  Sequence 13 from patent US 5683985.
ACCESSION   I72382
VERSION     I72382.1  GI:3008521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Chu,B.Chen.Fei. and Orgel,L.
  TITLE     Oligonucleotide decoys and methods relating thereto
  JOURNAL   Patent: US 5683985-A 13 04-NOV-1997;
FEATURES             Location/Qualifiers
     source          1. .46
                     /organism="unknown"
BASE COUNT        3 a      21 c     15 g      7 t
ORIGIN

  Query Match           81.8%;  Score 18;  DB 6;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+03;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||
Db         24 GATCGGGGCGGGGCGAGC 7


RESULT 30
AC144487/c
LOCUS       AC144487                181162 bp    DNA     linear   VRT 17-MAY-2003
DEFINITION  Gasterosteus aculeatus clone CH213-160O9, complete sequence.
ACCESSION   AC144487
VERSION     AC144487.1  GI:30089725
KEYWORDS    HTG.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
  ORGANISM  Gasterosteus aculeatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
            Gasterosteidae; Gasterosteus.
REFERENCE   1  (bases 1 to 181162)
  AUTHORS   Kingsley,D., Peichel,C., Colosimo,P., Shapiro,M., Marks,M.,
            Blackman,B., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
  TITLE     Genomic studies of vertebrate diversity
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181162)
  AUTHORS   Kingsley,D., Peichel,C., Colosimo,P., Shapiro,M., Marks,M.,
            Blackman,B., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2003) Stanford Human Genome Center, 975
            California Avenue, Palo Alto, CA 94304, USA
REFERENCE   3  (bases 1 to 181162)
  AUTHORS   Kingsley,D., Peichel,C., Colosimo,P., Shapiro,M., Marks,M.,
            Blackman,B., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2003) Stanford Human Genome Center, 975
            California Avenue, Palo Alto, CA 94304, USA
COMMENT     The sequence of the clone was established as a mapping and
            sequencing collaboration at the Stanford Genome Evolution Center,
            funded by the NIH Centers of Excellence in Genomic Science (CEGS)
            initiative (http://cegs.stanford.edu). The clone was isolated from
            the BAC library CHORI213 built by Pieter deJong in collaboration
            with the Stanford Genome Evolution Center
            (http://www.chori.org/bacpac/). Further details on the source fish
            and construction of the library can be found at
            http://cegs.stanford.edu/stickleback_BAC_library_link.jsp.
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
```

```
FEATURES             Location/Qualifiers
     source          1. .181162
                     /organism="Gasterosteus aculeatus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:69293"
                     /clone="CH213-160O9"
BASE COUNT    52530 a  39127 c  37839 g  51666 t
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 5;  Length 181162;
  Best Local Similarity   90.5%;  Pred. No. 1.6e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||| | |
Db     150119 ATTCGATCGGGGCGGGGGGGG 150099


RESULT 31
AC127795
LOCUS       AC127795               274015 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-101A9, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC127795
VERSION     AC127795.3  GI:25139837
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 274015)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 274015)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 274015)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23269472.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GZWV
                Center clone name: CH230-101A9
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 219062 bases at least Q40
                Consensus quality: 222791 bases at least Q30
                Consensus quality: 225601 bases at least Q20
                Estimated insert size: 225142; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9143: contig of 9143 bp in length
            *     9144     9243: gap of unknown length
            *     9244   272830: contig of 263587 bp in length
            *   272831   272930: gap of unknown length
            *   272931   274015: contig of 1085 bp in length.
FEATURES             Location/Qualifiers
     source          1. .274015
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-101A9"
     misc_feature    1. .1184
```

```
                     /note="wgs_contig"
     misc_feature    9244. .11119
                     /note="wgs_contig"
BASE COUNT    61069 a  49181 c  50666 g  66519 t  46580 others
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 2;  Length 274015;
  Best Local Similarity   90.5%;  Pred. No. 1.5e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||| |||||||||||| |||
Db     228660 ATTCCATCGGGGCGGGGGGAG 228680


RESULT 32
AC112802
LOCUS       AC112802               301660 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-35H17, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC112802
VERSION     AC112802.3  GI:23101275
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 301660)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 301660)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 301660)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 18, 2002 this sequence version replaced gi:21737447.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GPDT
                Center clone name: CH230-35H17
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 228218 bases at least Q40
                Consensus quality: 231561 bases at least Q30
                Consensus quality: 233804 bases at least Q20
                Estimated insert size: 253557; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   197738: contig of 197738 bp in length
            *    197739   197838: gap of unknown length
            *    197839   301660: contig of 103822 bp in length.
FEATURES             Location/Qualifiers
     source          1. .301660
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-35H17"
     misc_feature    1. .1034
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    5898. .6267
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH276513"
     misc_feature    6224. .6827
                     /note="clone_boundary
```

```
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH276515"
     misc_feature    197839. .199723
                     /note="wgs_end_extension
                     clone_end:Sp6"
BASE COUNT    67072 a  49216 c  49408 g  69772 t  66192 others
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 2;  Length 301660;
  Best Local Similarity   90.5%;  Pred. No. 1.5e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||| |||||||||||| |||
Db      15144 ATTCCATCGGGGCGGGGGGAG 15164


RESULT 33
BC016449
LOCUS       BC016449               1119 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus ring finger protein 5, mRNA (cDNA clone MGC:19320
            IMAGE:4191746), complete cds.
ACCESSION   BC016449
VERSION     BC016449.1  GI:16741215
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1119)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1119)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 24 Row: i Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, GenomeScan gene prediction, Similarity but not identity
            to protein.
FEATURES              Location/Qualifiers
     source           1. .1119
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="FVB/N"
                      /db_xref="taxon:10090"
                      /clone="MGC:19320 IMAGE:4191746"
                      /tissue_type="Salivary gland, 10 week old female mouse"
                      /clone_lib="NCI_CGAP_SG2"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     gene             1. .1119
                      /gene="Rnf5"
                      /note="synonyms: 2410131O05Rik, NG2"
                      /db_xref="LocusID:54197"
                      /db_xref="MGI:1860076"
     CDS              89. .631
                      /codon_start=1
                      /product="ring finger protein 5"
                      /protein_id="AAH16449.1"
                      /db_xref="GI:16741216"
                      /db_xref="LocusID:54197"
                      /translation="MAAAEEEDGGPEGPNRERGGASATFECNICLETAREAVVSVCGH
                      LYCWPCLHQWLETRPDRQECPVCKAGISREKVVPLYGRGSQKPQDPRLKTPPRPQGQR
                      PAPESRGGFQPPFGDAGGFHFSFGVGAFPPFGFFTTVFNAHEPFRRGAGVDLGQGHPASS
                      WQDSLFLFLAIFFFFWLLSI"
BASE COUNT      244 a    292 c    296 g    287 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 10;  Length 1119;
  Best Local Similarity   94.7%;  Pred. No. 3.3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        136 CGAGCGGGGCGGGGCGAGC 154


RESULT 34
HSA243936
LOCUS       HSA243936              1123 bp    mRNA    linear   PRI 03-AUG-1999
DEFINITION  Homo sapiens mRNA for G16 protein (G16 gene located in the class
            III region of the major histocompatibility complex).
ACCESSION   AJ243936
VERSION     AJ243936.1  GI:5578772
KEYWORDS    G16 gene; G16 protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kendall,E., Sargent,C.A. and Campbell,R.D.
  TITLE     Human major histocompatibility complex contains a new cluster of
            genes between the HLA-D and complement C4 loci
  JOURNAL   Nucleic Acids Res. 18 (24), 7251-7257 (1990)
  MEDLINE   91081311
   PUBMED   2259622
REFERENCE   2
  AUTHORS   Khanna,A.
  JOURNAL   Thesis (1993) University of Oxford, Department of Biochemistry. MRC
            Immunochemistry Unit
REFERENCE   3
  AUTHORS   Khanna,A. and Campbell,R.D.
```

```
  TITLE     Characterisation of a novel gene, G16, in the class III region of
            the human Major Histocompatibility Complex
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 1123)
  AUTHORS   Aguado,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-1999) Aguado B., HGMP Resource Centre, MRC,
            Genome Campus, Hinxton, Cambridge, CB10 1SB, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .1123
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /cell_line="U937"
                     /note="class III region of the major histocompatibility
                     complex"
     gene            1. .1123
                     /gene="G16"
     CDS             101. .643
                     /gene="G16"
                     /function="unknown"
                     /note="contains a Ring Finger"
                     /codon_start=1
                     /product="protein G16"
                     /protein_id="CAB51286.1"
                     /db_xref="GI:5578773"
                     /translation="MAAAEEEDGGPEGPNRERGGASATFECNICLETAREAVVSVCGH
                     LYCWPCLHQWLETRPDRQECPVCKAGISREKVVPLYGRGSQKPQDPRLKTPPRPQGQR
                     PAPESRGGFQPFGDAGGFHFSFGVGAFPFGPFTTVFNAHEPFRRGAGVDLGQGHPASS
                     WQDSLFLFLAIFFFFWLLSI"
     polyA_signal    110. .1105
                     /gene="G16"
                     /note="putative"
     misc_RNA        978. .982
                     /gene="G16"
                     /note="mRNA instability motif (AUUUA)"
     polyA_signal    1088. .1093
                     /gene="G16"
     polyA_signal    1094. .1099
                     /gene="G16"
                     /note="putative"
BASE COUNT      240 a    292 c    299 g    292 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 9;  Length 1123;
  Best Local Similarity   94.7%;  Pred. No. 3.3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        148 CGAGCGGGGCGGGGCGAGC 166


RESULT 35
AF335479
LOCUS       AF335479               7125 bp    DNA     linear   BCT 02-FEB-2002
DEFINITION  Agrobacterium sp. IP I-671 putative transposase gene, partial cds;
            hydantoin utilization gene cluster, complete sequence; and putative
            resolvase gene, partial cds.
ACCESSION   AF335479
VERSION     AF335479.1  GI:18478559
KEYWORDS    .
SOURCE      Agrobacterium sp. IP I-671
  ORGANISM  Agrobacterium sp. IP I-671
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
REFERENCE   1  (bases 1 to 7125)
  AUTHORS   Hils,M. and Altenbuchner,J.
  TITLE     Hydantoin utilization genes of Agrobacterium sp. IP I-671
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 7125)
  AUTHORS   Hils,M. and Altenbuchner,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2001) Institute of Industrial Genetics,
            University of Stuttgart, Allmandring 31, Stuttgart 70569, Germany
FEATURES             Location/Qualifiers
     source          1. .7125
                     /organism="Agrobacterium sp. IP I-671"
                     /mol_type="genomic DNA"
                     /strain="IP I-671"
                     /db_xref="taxon:173261"
     CDS             complement(<1. .445)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAL73197.1"
                     /db_xref="GI:18478560"
                     /translation="MAKRKLLKDQDRRKLVDIPVDEDSLIRHYSLSLADRLEIELRRR
                     NHNRLGFAIQLCLMRYPGRVLGAEEAPPPAMLRYVADQIGAAPESFALYARREETRRD
                     HTARLMVYLETRSAIAQDRRAALLAAIHAATMSDDGAAIVSSIVAA"
     gene            complement(576. .1286)
                     /gene="hyuN"
     CDS             complement(576. .1286)
                     /gene="hyuN"
                     /note="HyuN"
                     /codon_start=1
                     /transl_table=11
                     /product="putative NADPH dependent flavin oxidoreductase"
                     /protein_id="AAL73198.1"
                     /db_xref="GI:18478561"
                     /translation="MTSHTSTSAAAHDDCSVASAFDRRYGAPTFNQIPSADELVSDCI
                     ALMLKHRSVRRYTQDPLPAGTLEILMAAGQSAATSSNMQTVSVIAVTDPEMKARLART
                     CAGQDFIANAPLLLCFVTDLARPARIATTIGADLFALPMIDTFLASISDCSIFAQNVV
                     LAAESLGMGTCYVGSLRNRADLVSKELNVPSGSAVLFGLCIGYEHPERITNVRPRFPQ
                     KGVRTRAVRNRTLRPNGA"
     gene            complement(1311. .2684)
                     /gene="hyuH"
     CDS             complement(1311. .2684)
                     /gene="hyuH"
                     /note="HyuH"
                     /codon_start=1
                     /transl_table=11
                     /product="D-hydantoinase"
                     /protein_id="AAL73199.1"
                     /db_xref="GI:18478562"
                     /translation="MDIIIKNGTIVTADGISRADIGIKDGKIVQIGGALGPAERTIDA
                     SGRYVFPGGVDVHTHVETVSFNTCSADTFATATVAAACGGTTTIVDFCQQDRGRSLTD
                     AVAKWDGMAGGKSAIDYGYHIIVLDPTDSVIEELEVLPELGITSFKVFMAYRGMNMID
                     DVTLLKTLDKAARTGSLVMVHAENGDAADYLRNKFVAEGKTAPIYHALSRPPRIEAEA
                     TARALALAEIVDAPIYIVHVTCEESLDEVMRSKARGVHALAETCTHYLYLTKEDLERP
                     GFEGAKYVFTPPARAKKDHEILWNALRNGAFETVSSDHCSWLFRGHKDKGRNDFRAIP
                     NGAPGVEERLMMVYQGVNEGRLSLTQFVELVATRPAKVFGMFPEKGTIAVGSDADIVL
                     WDPEAEMVIEQSAMHNAMDYSTYEGRKVKGVPKTVLLRGRVIVEDGSYVGAPTDGQFL
                     KRRRYKQ"
     gene            3154. .4068
                     /gene="hyuC"
     CDS             3154. .4068
                     /gene="hyuC"
                     /note="HyuC"
                     /codon_start=1
                     /transl_table=11
                     /product="D-carbamoylase"
                     /protein_id="AAL73200.1"
                     /db_xref="GI:18478563"
                     /translation="MTRQMILAVGQQGPIARAETREQVIARLLDMLANAASRGVNFIV
                     FPELAVTTFFPRWHLTDEAELDSFYETEMPGPLTRPLFEKAAELGIGFNFGYAELVVE
                     GGVKRRFNTSILVDRSGKIIGKYRKVHLPGHKEYEAYRPFQHLEKRYFEPGDMGFPVY
                     DVDAAKMGMFICNDRRWPEAWRVMGLKGAEIICGGYNTPTHNPAVPQHDHLTSFHHLL
                     SMQAGSYQNGAWSAAAGKVGMEEDCMLLGHSCIVAPTGEIVALTTTLEDEVITAMIDL
                     DRCRELREHIFNFKAHRQPQHYGLIAEL"
     gene            4085. .5338
                     /gene="hyuD"
     CDS             4085. .5338
```

```
                     /gene="hyuD"
                     /note="HyuD"
                     /codon_start=1
                     /transl_table=11
                     /product="putative D-amino acid dehydrogenase"
                     /protein_id="AAL73201.1"
                     /db_xref="GI:18478564"
                     /translation="MRVVVLGAGVVGIATAWYLSKAGHDVTVIDREPAAGFGTSYGNA
                     GQISPALSSPWAFPGLVKKAVHWAFAEYPPLIVGRVPDAAMARFLWRMWRASSVDQYV
                     TAKRAMVTLAEYSRDCFRALRREIPIDYAGRQRGLLVAFRSPEQSDGYARDLAVLDEL
                     RVPYRRLSREEVAEVEPNIAPTSGVIGGVQLETDESGDCFRFTQALAARCAENGVRFV
                     YGHAASALLMENGRIKAVRTAEADVEMDAVVVALGVWSNALLKPLGIDLPIYPVKGYS
                     LTVKADPDTFGPMATISDETYKVGVTNLGDRIRVGGTAELAGFDVSQPEKRYAGLYQT
                     LKSLFPKVPEAAIADAERWSGLRPVTPDGPPLIGRSRIPNLYINSGHGTLGWTMSCGS
                     GRLLADVIGDERPAVSLSRFAPTAL"
     gene            5358. .6101
                     /gene="hyuA"
     CDS             5358. .6101
                     /gene="hyuA"
                     /note="HyuA"
                     /codon_start=1
                     /transl_table=11
                     /product="hydantoin-racemase"
                     /protein_id="AAL73202.1"
                     /db_xref="GI:18478565"
                     /translation="MKIKVINPNTTWTMTDKIAGAARAAAAPGTEIVAVSPDMGPVSI
                     EGYYDEVFAAVGVVDEVRKGELEGCDGYVIACFGDPGLNAAREVARGPVIGIAEAAMH
                     AASLIGGSFSIISMLGRSRGVLEHLVHSYGMAHKCRSVRMTDLPVLEFEEEGSDARRI
                     VVEECRRAIEQDHADSVLLGCGGMSDLMAYVSQEIGAPAIDGVSAGVKLVEALVGMGL
                     GTSKRAGLRLSDRESLHCSFSSFAPSRPR"
     CDS             6875. .>7125
                     /note="ORF7"
                     /codon_start=1
                     /transl_table=11
                     /product="putative resolvase"
                     /protein_id="AAL73203.1"
                     /db_xref="GI:18478566"
                     /translation="MLIGYMRVSSGDERQSVALQRDALLAAGVDQRHLHQDRASGARD
                     DRPGLKACLAELCEGDVLVVWKLDRLGRSLSHLIRIVEDP"
BASE COUNT     1403 a    2103 c    2146 g    1473 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 1;  Length 7125;
  Best Local Similarity   94.7%;  Pred. No. 2.9e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||||| ||||||||||||||
Db        902 CGATCCGGGCGGGGCGAGC 920


RESULT 36
AX346590
LOCUS       AX346590                9646 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1661 from Patent WO0200928.
ACCESSION   AX346590
VERSION     AX346590.1  GI:18494476
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 1661 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .9646
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     2765 a     150 c    2136 g    4595 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 6;  Length 9646;
  Best Local Similarity   94.7%;  Pred. No. 2.8e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||| |
Db       5203 TCGATCGGGGCGGGGCGGG 5221


RESULT 37
HSLUCA1/c
LOCUS       HSLUCA1                30377 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone XXcos-1 on chromosome 3, complete
            sequence.
ACCESSION   Z74618
VERSION     Z74618.1  GI:1405891
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 30377)
  AUTHORS   Wilkinson,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep XXcos-1 is part
            of a clone contig from chromosome 3p21.3 described in Ming-Hui Wei
            et al., Cancer Research 56, 1487-1492, 1996, and is from a
            Stratagene male, caucasian, placental cosmid library. VECTOR:
            pWE15.
FEATURES             Location/Qualifiers
     source          1. .30377
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="XXcos-1"
                     /clone_lib="SCLUCA"
BASE COUNT     7280 a    8596 c    7732 g    6769 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 9;  Length 30377;
  Best Local Similarity   94.7%;  Pred. No. 2.6e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db      29479 CGAGCGGGGCGGGGCGAGC 29461


RESULT 38
RN520N23/c
LOCUS       RN520N23              106278 bp    DNA     linear   HTG 04-JUN-2002
DEFINITION  Rattus norvegicus clone RPCI-31-520N23 strain Brown Norway, ***
            SEQUENCING IN PROGRESS ***, 138 unordered pieces.
ACCESSION   AL732653
VERSION     AL732653.1  GI:21326744
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Sudbrak,R., Borzym,K., Mueller,I., Klages,S., Kosiura,A.,
            Walter,L., Guenther,E., Hurt,P., Lehrach,H., Himmelbauer,H. and
            Reinhardt,R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 106278)
  AUTHORS   MOLGENR.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2002) MPIMG, Abt.Lehrach, Max Planck Institut
            Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany
COMMENT     contig 01          1. .753
            contig 02        854. .1659
            contig 03       1760. .2272
            contig 04       2373. .2868
            contig 05       2969. .3469
            contig 06       3570. .3886
            contig 07       3987. .4673
            contig 08       4774. .5688
            contig 09       5789. .6264
            contig 10       6365. .6870
            contig 11       6971. .7382
            contig 12       7483. .7903
            contig 13       8004. .8333
            contig 14       8434. .8903
            contig 15       9004. .9420
            contig 16       9521. .9680
            contig 17       9781. .10993
            contig 18      11094. .12178
            contig 19      12279. .13325
            contig 20      13426. .14231
            contig 21      14332. .14718
            contig 22      14819. .15502
            contig 23      15603. .15740
            contig 24      15841. .16159
            contig 25      16260. .16507
            contig 26      16608. .17067
            contig 27      17168. .17424
            contig 28      17525. .17641
            contig 29      17742. .18174
            contig 30      18275. .18414
            contig 31      18515. .18723
            contig 32      18824. .19217
            contig 33      19318. .19460
            contig 34      19561. .19667
            contig 35      19768. .20117
            contig 36      20218. .20654
            contig 37      20755. .20884
            contig 38      20985. .21387
            contig 39      21488. .21774
            contig 40      21875. .22052
            contig 41      22153. .22547
            contig 42      22648. .22955
            contig 43      23056. .23543
            contig 44      23644. .23773
            contig 45      23874. .24139
            contig 46      24240. .24602
            contig 47      24703. .25176
            contig 48      25277. .25673
            contig 49      25774. .26363
            contig 50      26464. .26897
            contig 51      26998. .29242
            contig 52      29343. .29710
            contig 53      29811. .29973
            contig 54      30074. .30247
            contig 55      30348. .31308
            contig 56      31409. .34048
            contig 57      34149. .35157
            contig 58      35258. .36473
            contig 59      36574. .37246
            contig 60      37347. .37850
            contig 61      37951. .38902
            contig 62      39003. .39823
            contig 63      39924. .40334
            contig 64      40435. .40887
            contig 65      40988. .41046
            contig 66      41147. .41464
            contig 67      41565. .42047
            contig 68      42148. .43221
            contig 69      43322. .44302
            contig 70      44403. .45038
            contig 71      45139. .45672
            contig 72      45773. .46403
            contig 73      46504. .47013
            contig 74      47114. .47546
            contig 75      47647. .47978
            contig 76      48079. .48523
            contig 77      48624. .49141
            contig 78      49242. .49811
            contig 79      49912. .51036
            contig 80      51137. .51679
            contig 81      51780. .52266
            contig 82      52367. .53022
            contig 83      53123. .53855
            contig 84      53956. .54255
            contig 85      54356. .54832
            contig 86      54933. .55626
            contig 87      55727. .56558
            contig 88      56659. .57428
            contig 89      57529. .58075
            contig 90      58176. .59033
            contig 91      59134. .59693
            contig 92      59794. .60166
            contig 93      60267. .60872
            contig 94      60973. .61737
            contig 95      61838. .62593
            contig 96      62694. .63459
            contig 97      63560. .64797
            contig 98      64898. .65488
            contig 99      65589. .67044
            contig 100      67145. .67584
            contig 101      67685. .68823
            contig 102      68924. .69802
            contig 103      69903. .70536
            contig 104      70637. .71214
            contig 105      71315. .71625
            contig 106      71726. .72513
            contig 107      72614. .74090
            contig 108      74191. .74968
            contig 109      75069. .76178
            contig 110      76279. .76723
            contig 111      76824. .78451
            contig 112      78552. .80226
            contig 113      80327. .81354
            contig 114      81455. .83535
            contig 115      83636. .85393
            contig 116      85494. .87577
            contig 117      87678. .90303
            contig 118      90404. .91635
```

```
          contig 119      91736. .92784
          contig 120      92885. .93370
          contig 121      93471. .93972
          contig 122      94073. .96680
          contig 123      96781. .97314
          contig 124      97415. .97765
          contig 125      97866. .98251
          contig 126      98352. .98406
          contig 127      98507. .98635
          contig 128      98736. .99395
          contig 129      99496. .100370
          contig 130     100471. .100855
          contig 131     100956. .101534
          contig 132     101635. .102482
          contig 133     102583. .103329
          contig 134     103430. .104188
          contig 135     104289. .104450
          contig 136     104551. .105037
          contig 137     105138. .105707
          contig 138     105808. .106278.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 138 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1       753: contig of 753 bp in length
          *      754       853: gap of 100 bp
          *      854      1659: contig of 806 bp in length
          *     1660      1759: gap of 100 bp
          *     1760      2272: contig of 513 bp in length
          *     2273      2372: gap of 100 bp
          *     2373      2868: contig of 496 bp in length
          *     2869      2968: gap of 100 bp
          *     2969      3469: contig of 501 bp in length
          *     3470      3569: gap of 100 bp
          *     3570      3886: contig of 317 bp in length
          *     3887      3986: gap of 100 bp
          *     3987      4673: contig of 687 bp in length
          *     4674      4773: gap of 100 bp
          *     4774      5688: contig of 915 bp in length
          *     5689      5788: gap of 100 bp
          *     5789      6264: contig of 476 bp in length
          *     6265      6364: gap of 100 bp
          *     6365      6870: contig of 506 bp in length
          *     6871      6970: gap of 100 bp
          *     6971      7382: contig of 412 bp in length
          *     7383      7482: gap of 100 bp
          *     7483      7903: contig of 421 bp in length
          *     7904      8003: gap of 100 bp
          *     8004      8333: contig of 330 bp in length
          *     8334      8433: gap of 100 bp
          *     8434      8903: contig of 470 bp in length
          *     8904      9003: gap of 100 bp
          *     9004      9420: contig of 417 bp in length
          *     9421      9520: gap of 100 bp
          *     9521      9680: contig of 160 bp in length
          *     9681      9780: gap of 100 bp
          *     9781     10993: contig of 1213 bp in length

  Query Match             79.1%;  Score 17.4;  DB 2;  Length 106278;
  Best Local Similarity   94.7%;  Pred. No. 2.4e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| |||||||||||||||
Db      81934 CGAGCGGGGCGGGGCGAGC 81916


RESULT 39
AC132215
LOCUS       AC132215               126323 bp    DNA     linear   PLN 03-SEP-2002
DEFINITION  Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OSJNBa0076E06, from chromosome 3, complete sequence.
ACCESSION   AC132215
VERSION     AC132215.1  GI:22657512
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 126323)
  AUTHORS   McCombie,W.R., de la Bastide,M., Spiegel,L., Preston,R.,
            Nascimento,L., Zutavern,T., Balija,V., Bell,M., Miller,B.,
            Katzenberger,F., Muller,S., Sullivan,P., Yang,C., Dike,S.,
            O'Shaughnessy,A., Palmer,L. and Dedhia,N.
  TITLE     Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OSJNBa0076E06, from chromosome 3, complete sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 126323)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2002) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.  The nucleotide
            sequence of this BAC clone was generated by combining Syngenta,
            Monsanto and Cold Spring Harbor Laboratory Genome Center sequencing
            data.
FEATURES             Location/Qualifiers
     source          1. .126323
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="3"
                     /clone="OSJNBa0076E06"
                     /clone_lib="HindIII"
     misc_feature    4490. .4506
                     /note="We believe the assembly to be correct.  The
                     sequence is a mononucleotide (C) repeat in which the exact
                     number of Cs is unknown.  The majority of subclones in
                     the area agree with the assembly, however, two subclones
                     show two additional Cs."
     misc_feature    91540. .91640
                     /note="We believe the assembly to be correct.  The
                     sequence is covered by a PCR product which was amplified
                     with a high fidelity polymerase.  The sequence is of high
                     quality and there is partial coverage by several subclones
                     with quality below phred30."
BASE COUNT    35405 a  27507 c  26103 g  37308 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 8;  Length 126323;
  Best Local Similarity   94.7%;  Pred. No. 2.4e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||| |
Db       4300 TCGATCGGGGCGGGGCGGG 4318


RESULT 40
AC137547
LOCUS       AC137547               144969 bp    DNA     linear   PLN 02-APR-2003
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 3 clone
            OSJNBa0052J20, complete sequence.
```

```
ACCESSION   AC137547
VERSION     AC137547.2  GI:29469501
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 144969)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Kim,H.-R., Rambo,T., Currie,J. and
            Collura,K.
  TITLE     Rice Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 144969)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Kim,H.-R., Rambo,T., Saski,C.,
            Currie,J. and Collura,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2002) Arizona Genomics Institute, University of
            Arizona, 303 Forbes, Tucson, AZ 85721, USA
REFERENCE   3  (bases 1 to 144969)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Kim,H.-R., Rambo,T., Currie,J. and
            Collura,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2003) Arizona Genomics Institute, University of
            Arizona, 303 Forbes, Tucson, AZ 85721, USA
COMMENT     On Apr 2, 2003 this sequence version replaced gi:25188927.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality
            >=30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by more than
            one plasmid subclone ; and the assembly was confirmed by
            restriction digest.There is a TA dinucleotide repeat ranging from
            10-30 pairs at 92656-91715.From base 10472-10575 the subclone reads
            for OSJNBa0052J20 have deleted sequence but the PCR done off this
            BAC DNA matches the overlapping clone OSJNBb0014I10.  There are
            Bacterial Transposonsat the following locations: 28628-33998,
            45024-48063 and 43861-47164.  At bases 44656-44749 only transposon
            reads cover this area.  There is a repeat of TAA (36-38 repeats) at
            base 50810.  The assembly overlaps from base 1-33264 with
            OSJNBb0014I10 (accession #AC126222). The overlap is from bases
            80683-113946 on OSJNBb0014I10.  The assembly overlaps from base
            127973-144969 with OSJNBa0076E06 (accession #AC132215). The overlap
            is from bases 1-16998 on OSJNBa0076E06. The nucleotide sequence of
            this BAC clone was generated by combining Syngenta, Monsanto and
            Arizona Genomics Institute sequencing data.
FEATURES             Location/Qualifiers
     source          1. .144969
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39947"
                     /chromosome="3"
                     /clone="OSJNBa0052J20"
BASE COUNT    40242 a  32367 c  31769 g  40591 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 8;  Length 144969;
  Best Local Similarity   94.7%;  Pred. No. 2.4e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
              ||||||||||||||||| |
Db     132271 TCGATCGGGGCGGGGCGGG 132289


Search completed: February 18, 2004, 16:42:29
Job time : 1543 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 16:09:36 ; Search time 172 Seconds
(without alignments)
345.277 Million cell updates/sec

Title: US-10-026-341A-2
Perfect score: 22
Sequence: 1 attcgatcggggcggggcgagc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	22	100.0	22	17	AAT18820	SP1 motif. Homo s
	2	22	100.0	22	18	AAT77128	Sp1 consensus. Sy
	3	22	100.0	22	20	AAX76050	cAMP response elem
	4	22	100.0	22	21	AAD01054	Oligonucleotide wi
c	5	22	100.0	22	21	AAA46366	Nucleotide sequenc
	6	22	100.0	22	22	AAD10308	SP1 oligonucleotid
	7	22	100.0	22	22	AAD06435	PCR primer Sp1, us
	8	22	100.0	22	24	AAD44321	Decoy Sp1 oligonuc
	9	22	100.0	22	24	AAD40710	Mouse osterix bind
	10	22	100.0	22	24	ABQ74742	Glucose carrier ty
	11	22	100.0	22	24	ABA92271	Sp-1 wild-type oli
	12	22	100.0	22	24	ABA93509	Regulatory element
	13	22	100.0	22	25	ABZ58136	Transcription fact
	14	22	100.0	49	22	AAF61841	SP1-specific DNA p
c	15	22	100.0	49	22	AAF61841	SP1-specific DNA p
	16	22	100.0	49	22	AAF61843	SP1-specific DNA p
c	17	22	100.0	49	22	AAF61843	SP1-specific DNA p
	18	21	95.5	21	22	AAF99620	Immunostimulatory
c	19	21	95.5	21	22	AAF99621	Immunostimulatory
	20	21	95.5	21	24	ABS78341	Angiogenesis inhib
c	21	21	95.5	21	24	ABS78342	Angiogenesis inhib
	22	21	95.5	21	24	ABL38746	Immunostimulatory
c	23	21	95.5	21	24	ABL38839	Immunostimulatory
	24	21	95.5	21	25	ABT17254	Transcription fact
c	25	21	95.5	21	25	ABT17255	Transcription fact
c	26	21	95.5	63	25	ABT17322	Transcription fact
	27	20.4	92.7	22	24	ABK89728	Oestrogen response
	28	20	90.9	20	19	AAV46004	Immune adjuvant SP
	29	20	90.9	20	21	AAZ99627	Nucleotide sequenc
	30	20	90.9	20	24	AAL39229	Murine Toll-like r
	31	18.8	85.5	22	24	ABA92272	Sp-1 mutant oligon
	32	18.4	83.6	22	15	AAQ67304	Detection probe fo
	33	18	81.8	45	13	AAQ30483	Oligonucleotide co
c	34	18	81.8	45	13	AAQ30483	Oligonucleotide co
c	35	18	81.8	47	13	AAQ30484	Oligonucleotide co
c	36	17.4	79.1	183	25	ABA00746	Enh1 enhancer. Sy
	37	17.4	79.1	722	24	ABQ44940	Oligonucleotide fo
c	38	17.4	79.1	722	24	ABQ44941	Oligonucleotide fo
	39	17.4	79.1	747	24	ABQ43520	Oligonucleotide fo
c	40	17.4	79.1	747	24	ABQ43521	Oligonucleotide fo
	41	17.4	79.1	9646	24	ABL33688	Human immune syste
c	42	17.2	78.2	10035	25	ABZ66813	Orthosomycin biosy
c	43	17.2	78.2	11115	23	ABL50562	Micromonospora car
	44	17	77.3	540	24	ABQ23090	Oligonucleotide fo
c	45	17	77.3	540	24	ABQ23091	Oligonucleotide fo

ALIGNMENTS

RESULT 1
AAT18820
ID AAT18820 standard; DNA; 22 BP.
XX
AC AAT18820;
XX
DT 17-AUG-1996 (first entry)
XX
DE SP1 motif.
XX
KW NF-X1; transcription factor; major histocompatibility complex; MHC;
KW allergy; HLA-DRA; ds.
XX
OS Homo sapiens.
XX
PN WO9612823-A1.
XX
PD 02-MAY-1996.
XX
PF 20-OCT-1995; 95WO-US12749.
XX
PR 21-OCT-1994; 94US-0327832.
XX
PA (HARD) HARVARD COLLEGE.
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Ono SJ, Strominger JL;
XX
DR WPI; 1996-230621/23.
XX
PT Transcription factor, NF-X1 and DNA encoding it - used in regulation

PT of MHC class II expression and in treatment of allergic disease
XX
PS Example 4; Page 42; 93pp; English.
XX
CC Recombinant transcription factor NF-X1 (see AAR94957) forms a
CC specific complex with the HLA-DRA X1 box oligonucleotide (AAT18817)
CC which is competed for by 100-fold excess cold, double-stranded
CC oligonucleotides containing the analagous regions from the HLA-DRB,
CC -DPA, -DPB, -DQA and -DQB promoters, but not by HLA-DRA Y-box
CC (AAT18818), S-box (AAT18819), SP1 (AAT18820) or the interferon-beta gene
CC positive-regulatory domain II element (AAT18821). It is concluded
CC that NF-X1 binds sequence-specifically with all human class II
CC major histocompatibility X1 boxes (see also AAT18812).
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 17; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 2
AAT77128
ID AAT77128 standard; DNA; 22 BP.
XX
AC AAT77128;
XX
DT 07-DEC-1997 (first entry)
XX
DE Sp1 consensus.
XX
KW 17-Beta-hydroxysteroid dehydrogenase type I; HSD17B1; human;
KW promoter; ds.
XX
OS Synthetic.
XX
PN WO9720942-A1.
XX
PD 12-JUN-1997.
XX
PF 04-DEC-1996; 96WO-FI00647.
XX
PR 05-DEC-1995; 95US-0007976.
XX
PA (OIKA/) OIKARINEN J A.
PA (PELT/) PELTOKETO E H.
PA (PIAO/) PIAO Y.
PA (VIHK/) VIHKO R K.
XX
PI Oikarinen JA, Peltoketo EH, Piao Y, Vihko RK;
XX
DR WPI; 1997-319788/29.
XX
PT Human 17-beta-hydroxysteroid dehydrogenase type I (HSD17B1)
PT transcription regulatory elements - used for identifying agents
PT which can up- or down-regulate HSD17B1 expression to increase or
PT decrease oestrogen production
XX
PS Example 6; Page 38; 69pp; English.
XX
CC This oligonucleotide comprises a consensus sequence for Sp1
CC binding sites. It was used with oligonucleotides (see
CC AAT77122-24 and AAT77126-27) based on the promoter region of the human
CC 17-beta-hydroxysteroid dehydrogenase type I (HSD17B1) gene (see
CC AAT77112), and with an AP-2 consensus oligonucleotide (see AAT77125),
CC in the detailed characterisation of the HSD17B1 promoter, and to
CC examine the role of Sp1 and AP-2 binding sites in promoter
CC function.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 18; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 3
AAX76050
ID AAX76050 standard; DNA; 22 BP.
XX
AC AAX76050;
XX
DT 30-JUL-1999 (first entry)
XX
DE cAMP response element oligonucleotide SEQ ID NO:18.
XX
KW CRE; cAMP response element; transcription factor decoy; cis-element;
KW tumour growth inhibitor; palindrome; hairpin; cancer; metabolism;
KW gene transcription regulation; inhibiting proliferation; ds.
XX
OS Synthetic.
XX
PN WO9926634-A1.
XX
PD 03-JUN-1999.
XX
PF 23-NOV-1998; 98WO-US25307.
XX
PR 24-NOV-1997; 97US-0977643.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Cho-Chung YS;
XX
DR WPI; 1999-347612/29.
XX
PT Nucleic acids that compete with response elements for transcription
PT factors
XX
PS Example 10; Page 54; 83pp; English.
XX
CC The present invention describes a composition (A) comprising one or more
CC nucleic acids (I) that compete with cAMP (cyclic adenosine monophosphate)
CC response element (CRE) enhancer DNA for binding to transcription factors
CC (TF). (I) are used to regulate gene transcription in cells, in vitro or
CC in vivo, specifically for inhibiting proliferation of cancer cells, but
CC possibly also for regulation of metabolism in hepatitis B and other
CC viruses. HCT-15 human multidrug resistant colon carcinoma cells (2
CC million) were inoculated subcutaneously into the flank of nude mice,
CC then the CRE oligonucleotide 5'-TGACGTTCATGACGTTCATGACGTTCA-3' injected
CC intraperitoneally at doses of 0.1 mg, 5 times per week, once the tumour
CC had reached 30-50 mg. This treatment resulted in over 85% reduction in
CC tumour growth, relative to an untreated control. (I) have high affinity
CC for TF and can inhibit growth of cancer cells without adverse effects on
CC normal cells (contrast use of antisense RNA). The method does not
CC require knowledge of the target gene sequence, only of the response
CC element sequence. The present sequence is used in the exemplification
CC of the present invention.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 20; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 4
AAD01054
ID AAD01054 standard; DNA; 22 BP.
XX
AC AAD01054;
XX
DT 04-OCT-2000 (first entry)
XX
DE Oligonucleotide with Sp1 consensus site for EMSA of human cyclin A1 gene.
XX
KW Human; cyclin A1; transgenic; stem cell; male germ cell; promoter;
KW fluorescent protein; light emitting reporter protein; genetic defect;
KW gene therapy; molecular engineering; trauma; vertebrate; Sp1;
KW electrophoretic mobility shift assay; EMSA; ds.
XX
OS Unidentified.
XX
PN WO200029602-A1.
XX
PD 25-MAY-2000.
XX
PF 15-APR-1999; 99WO-US08277.
XX
PR 13-NOV-1998; 98US-0191920.
PR 13-NOV-1998; 98WO-US24238.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
PA (UNLO) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
PI Readhead CW, Winston R, Koeffler HP, Muller C;
XX
DR WPI; 2000-387812/33.
XX
PT Producing selectable transgenic stem cells for use in molecular
PT engineering, e.g. creating transgenic animals -
XX
PS Example 12; Page 41; 98pp; English.
XX
CC The patent discloses in vivo and in vitro methods for producing
CC selectable transgenic stem cells by transfecting male germ cells with a
CC transfection mixture comprising a nucleic acid construct containing a
CC transcriptional unit of a stem cell-specific promoter, e.g. cyclin
CC A1 promoter, operably linked to a gene encoding a fluorescent or
CC light emitting protein. These methods may be used for producing
CC transgenic vertebrate animals and for repairing genetic defects.
CC They are suitable for germ line and stem cell line gene therapy.
CC They are also valuable in identifying cell lineages before full
CC differentiation and to facilitate modification and/or engineering of
CC specific tissues in vitro for their subsequent transplantation in the
CC treatment of disease or trauma. The present sequence is an
CC oligonucleotide, containing Sp1 consensus site, used in electrophoretic
CC mobility shift assay using nuclear extract from Hela cells to
CC analyse human cyclin A1 gene. Cyclin A1 promoter derived from
CC the 5' end of cyclin A1 gene is stem cell-specific and used in a
CC DNA construct for producing selectable transgenic stem cells.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22
```

RESULT 5
AAA46366/c
ID AAA46366 standard; DNA; 22 BP.
XX
AC AAA46366;
XX
DT 04-SEP-2000 (first entry)
XX
DE Nucleotide sequence of a SP1 element oligonucleotide.
XX
KW Malignant cell; OP2 C2-6; OP7 C3-1; OP9 A4-2; OP11 C1-3; OP11 G2-10;
KW FAS OP13; FAS OP17; CINN 1; CINN 2; anti-neoplastic drug; malignancy;
KW cancer; brain cancer; intracranial neoplasm; ss.
XX
OS Unidentified.
XX
PN WO200028090-A2.
XX
PD 18-MAY-2000.
XX
PF 12-NOV-1999; 99WO-US27005.
XX
PR 12-NOV-1998; 98US-0108120.
PR 27-JUL-1999; 99US-0145640.
XX
PA (NYXI-) NYXIS INC.
XX
PI Kroes RA, Moskal JR, Jastrow A, Yamamoto H;
XX
DR WPI; 2000-376595/32.
XX
PT Method for ascertaining the propensity of a cell for malignant
PT phenotype, used for diagnosis of cancers, including those of the lung,
PT breast and brain -
XX
PS Disclosure; Page 44; 66pp; English.
XX
CC The specification describes a method for ascertaining the propensity of
CC a cell for malignant phenotype. The method comprises assaying a cell or
CC biological sample for a signal indicating the transcription of a nucleic
CC acid transcript. The transcript is from at least one gene selected from
CC the genes encoded for by or containing the characteristic nucleic acid
CC sequences of OP2 C2-6, OP7 C3-1, OP9 A4-2, OP11 C1-3, OP11 G2-10,
CC FAS OP13, FAS OP17, CINN 1 or CINN 2 (AAA46354-62, respectively). Cells
CC identified using the method can be used for ascertaining the suitability
CC of an anti-neoplastic drug candidate for efficacy in treating a
CC malignancy. The method can be used in the diagnosis of many types of
CC cancers, including cancers of the breast, prostate, colon and lung. It
CC may also be used for diagnosis and treatment of brain cancers, including
CC intracranial neoplasms such as those of the skull, meninges, cranial
CC nerves, neuroglia, ependyma, pituitary or pineal body, as well as those
CC of metastatic origin. The present sequence represents a SP1 element
CC oligonucleotide, which is used in the course of the invention.
XX
SQ Sequence 22 BP; 3 A; 11 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db         22 ATTCGATCGGGGCGGGGCGAGC 1
```

RESULT 6
AAD10308
ID AAD10308 standard; DNA; 22 BP.
XX
AC AAD10308;
XX
DT 24-SEP-2001 (first entry)
XX
DE SP1 oligonucleotide to determine nuclear factor kappa B activation.
XX
KW Human; Par-4; presenilin; PS; neuroprotective; nuclear factor kappa B;
KW NF-kappa B; neuronal degeneration; spinal muscular atrophy; paralysis;

KW peripheral neuropathy; motorneuron disorder; neurodegenerative disorder;
KW Parkinson's disease; Meniere's disease; multiple sclerosis; Bell's palsy;
KW Huntington's chorea; Down's syndrome; amyotrophic lateral sclerosis; ALS;
KW nerve deafness; Alzheimer's disease; epilepsy; ds.
XX
OS Unidentified.
XX
PN WO200151671-A2.
XX
PD 19-JUL-2001.
XX
PF 08-JAN-2001; 2001WO-US00526.
XX
PR 10-JAN-2000; 2000US-0175200.
PR 04-JAN-2001; 2001US-0754949.
XX
PA (SCIO-) SCIOS INC.
XX
PI Mccarthy J, Cordell B;
XX
DR WPI; 2001-451872/48.
XX
PT Identifying inhibitors of neuronal degeneration useful for treating
PT e.g. Alzheimer's disease, by determining the ability of a compound to
PT induce nuclear factor kappa B activation, with the involvement of
PT presenilin or Par-4 -
XX
PS Example 5; Page 25; 66pp; English.
XX
CC The invention relates to human Par-4 protein, presenilin protein (PS1
CC and PS2) and their corresponding DNA molecules. The invention also
CC relates to a method for identifying inhibitors of neuronal degeneration,
CC comprising cotransfecting eukaryotic host cells expressing presenilin
CC (PS), with a Par-4 DNA, and an NF-kappa B dependent reporter construct,
CC exposing the cotransfected cells to a candidate molecule and monitoring
CC the ability of the candidate molecule to induce NF-kappa B activation.
CC Presenilin proteins participates in nuclear factor kappa B (NF-kappa B)
CC signalling and activation. The inhibitors of neuronal degeneration
CC are useful for treating neurodegenerative disorders such as Alzheimer's
CC disease, Parkinson's disease, epilepsy, multiple sclerosis, Huntington's
CC chorea, Down's syndrome, nerve deafness, Meniere's disease and also for
CC treating peripheral neuropathies, motorneuron disorders such as
CC amyotrophic lateral sclerosis (ALS), Bell's palsy and various conditions
CC involving spinal muscular atrophy and paralysis. The present DNA sequence
CC is SP-1 oligonucleotide which is used to determine nuclear factor kappa B
CC (NF-kappaB) activation.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 7
AAD06435
ID AAD06435 standard; DNA; 22 BP.
XX
AC AAD06435;
XX
DT 10-AUG-2001 (first entry)
XX
DE PCR primer Sp1, used in electrophoretic mobility shift assay.
XX
KW Human; Mcl-1 gene regulatory element; Mcl-1s/deltaTM variant;
KW neuronal cell; tumour cell; apoptosis; therapy; cancer; psoriasis;
KW diabetic retinopathy; corneal graft neovascularisation;
KW neovascular glaucoma; epithelial condition; autoimmune disease;
KW rheumatoid arthritis; systemic lupus erythematosus;
KW neurodegenerative disease; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200136594-A1.
XX
PD 25-MAY-2001.
XX
PF 14-JAN-2000; 2000WO-US00969.
XX
PR 16-NOV-1999; 99US-0166113.
XX
PA (DART-) DARTMOUTH COLLEGE.
XX
PI Craig RW, Bingle CD, Whyte M;
XX
DR WPI; 2001-343812/36.
XX
PT Novel Mcl-1 gene regulatory elements, useful for modulating expression
PT of Mcl-1 polypeptide or its variant which regulate apoptosis in
PT neuronal or tumor cells -
XX
PS Example 1; Page 63; 125pp; English.
XX
CC The present invention relates to Mcl-1 gene regulatory elements and the
CC variant Mcl-1s/deltaTM. The anti-apoptotic Mcl-1 protein is encoded by
CC exons 1, 2 and 3. The pro-apoptotic Mcl-1s/DeltaTM variant encoded by
CC exons 1 and 3 is obtained due to alternative mRNA splicing. The Mcl-1
CC gene regulatory element is useful for modulating the Mcl-1 gene
CC expression in a cell e.g., neuronal cell or tumour cell, such that
CC apoptosis of the cell is induced or cell viability is increased. The
CC Mcl-1 and its regulatory elements are used for treating pathological
CC conditions which include cancer, diabetic retinopathy, corneal graft
CC neovascularisation and neovascular glaucoma, epithelial conditions such
CC as psoriasis, autoimmune diseases like rheumatoid arthritis, systemic
CC lupus erythematosus, and neurodegenerative diseases. The present sequence
CC is a PCR primer Sp1, used in the electrophoretic mobility shift assay,
CC which is related to the invention.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 8
AAD44321
ID AAD44321 standard; DNA; 22 BP.
XX
AC AAD44321;
XX
DT 13-DEC-2002 (first entry)
XX
DE Decoy Sp1 oligonucleotide.
XX
KW Fibrotic condition; gene expression; cirrhosis; hypertrophic scar;
KW gene therapy; fibrosis; skin disorder; sclerodermic lesion; keloid;
KW trauma; surgery; Sp1; ds.
XX
OS Unidentified.
XX
PN WO200266071-A2.
XX
PD 29-AUG-2002.
XX
PF 21-DEC-2001; 2001WO-US49141.
XX
PR 03-JAN-2001; 2001US-259585P.

XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI Mauviel A;
XX
DR WPI; 2002-667041/71.
XX
PT Treating a fibrotic condition, e.g. cirrhosis, comprises administering
PT antisense Sp1 or decoy Sp1 oligonucleotides that inhibit transcription
PT or gene expression of an extra-cellular matrix gene or transforming
PT growth factor-beta -
XX
PS Claim 5; Page 38; 38pp; English.
XX
CC The invention relates to a method of treating a fibrotic condition (the
CC transcription or gene expression of an extra-cellular matrix (ECM) gene
CC or transforming growth factor-beta (TGF-beta) is inhibited in a mammal).
CC The method involves administering an antisense Sp1 or a decoy Sp1
CC oligonucleotide. The method, antisense Sp1 and decoy Sp1 oligonucleotide
CC are useful for treating fibrotic conditions e.g. cirrhosis, radiation
CC induced fibrosis, skin disorders (sclerodermic lesions, hypertrophic
CC scars, keloids), kidney fibrosis, lung fibrosis and myelofibrosis. They
CC are also useful for treating or preventing fibrotic scarring following
CC trauma or surgery. The invention is useful in gene therapy. The present
CC sequence is decoy Sp1 oligonucleotide..
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 24; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 9
AAD40710
ID AAD40710 standard; DNA; 22 BP.
XX
AC AAD40710;
XX
DT 30-OCT-2002 (first entry)
XX
DE Mouse osterix binding oligonucleotide, Sp1.
XX
KW Bone formation; transcription factor; osteoblast; Paget's disease;
KW osterix protein; glucocorticoid; osteoporosis; cytostatic; cytokine;
KW periodontal disease; tooth loss; bone fracture; rheumatoid arthritis;
KW metastatic bone disease; gene therapy; growth factor; osteopathic;
KW differentiation; mouse; ss.
XX
OS Mus musculus.
XX
PN WO200244380-A2.
XX
PD 06-JUN-2002.
XX
PF 30-NOV-2001; 2001WO-US44898.
XX
PR 30-NOV-2000; 2000US-0734329.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI De Crombrugghe B, Nakashima K, Zhou X;
XX
DR WPI; 2002-519587/55.
XX
PT Novel DNA segment encoding Osterix polypeptide which is a master
PT transcription factor that controls osteoblast differentiation and is
PT useful for treating osteoporosis, in patient by stimulating bone
PT formation -
XX
PS Example 4; Page 107; 144pp; English.
XX
CC The invention relates to a master bone formation transcription factor
CC that controls osteoblast differentiation, osterix protein and its
CC corresponding nucleic acid sequence. Osterix protein and its DNA and
CC agents that interact with the protein to activate or stimulate the
CC differentiation of bone cells is used for treatment of glucocorticoid
CC induced osteoporosis, Paget's disease, periodontal disease, tooth loss,
CC bone fractures, rheumatoid arthritis, metastatic bone disease, etc.
CC Osterix DNA is useful as probes or primers in nucleic acid hybridisation
CC experiments. It is also used in gene therapy. Osterix protein is useful
CC for controlling bone formation, serves as receptors of soluble
CC molecules, e.g. cytokines, growth factors, etc, as homing/adhesion/
CC rolling receptors mediating the migration of osteoblast, as signalling
CC receptors, thereby regulating function of osteoblasts, and/or ligands
CC of signalling receptors or osteoblast. It is also useful for purifying
CC osteoblasts, from a test composition suspected of containing the cells.
CC The present sequence is mouse osterix gene binding oligonucleotide used
CC in the exemplification of the invention.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 24; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 10
ABQ74742
ID ABQ74742 standard; DNA; 22 BP.
XX
AC ABQ74742;
XX
DT 24-OCT-2002 (first entry)
XX
DE Glucose carrier type 2 promoter related oligonucleotide #8.
XX
KW Hepatocyte nuclear factor; HNF; gene therapy; liver; pancreas;
KW glucose carrier type 2 promoter; tissue-specific expression;
KW glucose metabolism related disease; diabetes; ss.
XX
OS Synthetic.
XX
PN KR2001109882-A.
XX
PD 12-DEC-2001.
XX
PF 03-JUN-2000; 2000KR-0030606.
XX
PR 03-JUN-2000; 2000KR-0030606.
XX
PA (GENE-) GENEPIA CO LTD.
XX
PI Ahn YH, Cha JY, Huh MU, Kim GS, Kim HI;
XX
DR WPI; 2002-398926/43.
XX
PT DNA sequence for tissue-specific expression in liver or pancreas,
PT useful for gene therapy of glucose metabolism related disease, e.g.
PT diabetes -
XX
PS Example; Page 6; 15pp; Korean.
XX
CC The present invention describes a liver or pancreas tissue-specific
CC expression related DNA sequence (I). (I) is particularly a hepatocyte
CC nuclear factor (HNF) binding region present in glucose carrier type 2
CC promoter. (I) participates in tissue-specific expression in liver or
CC pancreas and HNF1 and HNF3 characteristically bind to it. (I) can be

CC used in a recombinant vector useful for gene therapy in glucose
CC metabolism related diseases, such as diabetes. The present sequence
CC represents an oligonucleotide which is used in the exemplification of
CC the present invention.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 24; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 11
ABA92271
ID ABA92271 standard; DNA; 22 BP.
XX
AC ABA92271;
XX
DT 10-JUN-2002 (first entry)
XX
DE Sp-1 wild-type oligonucleotide, used in EMSA.
XX
KW SP-1; neuron; antialzheimers; antiparkinsonian; antisclerotic;
KW neuroprotective; nootropic; anticonvulsant; vascular; hypotensive;
KW cerebroprotective; virucide; anti-HIV; diagnosis; therapy;
KW electrophoretic mobility shift assay; EMSA; ds.
XX
OS Homo sapiens.
XX
PN WO200215912-A1.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26527.
XX
PR 25-AUG-2000; 2000US-228201P.
PR 26-OCT-2000; 2000US-243295P.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Ratan RR, Chatterjee S;
XX
DR WPI; 2002-242023/29.
XX
PT Diagnosing and treating diseases associated with oxidative stress, DNA
PT damage or growth factor depletion, e.g. Alzheimer's and Parkinson's, by
PT administering e.g. mithramycin, chromomycin, daunomycin, olivomycin and
PT WP631 -
XX
PS Example 1; Page 23; 69pp; English.
XX
CC The present sequence is that of a double-stranded wild-type Sp-1
CC oligonucleotide, which was used in an electrophoretic mobility
CC shift assay to determine the effect of oxidative stress on Sp-1 DNA
CC binding, and to determine the effects of candidate compounds on Sp-1
CC protein levels. SP-1 DNA binding activity in cortical neurons was
CC shown to be low, but was dramatically enhanced by oxidative stress.
CC The invention provides methods for detecting and treating diseases
CC associated with oxidative stress, DNA damage or growth factor
CC depletion, and identifying agents for the treatment of such
CC conditions. A compound is deemed to be an inhibitor of oxidative
CC stress, DNA damage, growth factor depletion or cell death if it
CC reduces the protein level of an Sp family member or if it decreases
CC the binding of an Sp family member to DNA. A method for preventing
CC or treating a disease or disorder of the nervous system, the ageing
CC process or associated with apoptosis involves administering a
CC compound that inhibits the induction of an Sp family member or the
CC binding of an Sp family member to DNA, e.g. mithramycin,
CC chromomycin, daunomycin, olivomycin or WP631. Diseases and
CC disorders that can be treated include Alzheimer's disease,
CC Creutzfeldt-Jacob disease, kuru, Huntington's disease, aneurysm,
CC stroke associated with an increase in blood pressure, spinal cord
CC disease, spinal cord injury, brain injury, multiple system atrophy,
CC amylotrophic lateral sclerosis, progressive supranuclear palsy,
CC neurodegeneration associated with the ageing process, mitochondrial
CC disease, HIV infection, herpes infection and multiple sclerosis
CC (all claimed).
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 24; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 12
ABA93509
ID ABA93509 standard; DNA; 22 BP.
XX
AC ABA93509;
XX
DT 24-APR-2002 (first entry)
XX
DE Regulatory element related oligonucleotide probe SEQ ID NO:9.
XX
KW Pancreatic langerhans beta cell proliferation promoter; glucocorticoid;
KW apoptosis inhibitor; regulatory element; HGF; hepatocyte growth factor;
KW antidiabetic; apoptotic; diabetes; probe; ss.
XX
OS Synthetic.
XX
PN WO200193899-A1.
XX
PD 13-DEC-2001.
XX
PF 01-JUN-2001; 2001WO-JP04660.
XX
PR 02-JUN-2000; 2000JP-0170447.
PR 28-FEB-2001; 2001JP-0054072.
XX
PA (OKAM/) OKAMOTO H.
XX
PI Okamoto H;
XX
DR WPI; 2002-147646/19.
XX
PT Promoters for pancreatic Langerhans beta cell proliferation and
PT apoptosis inhibitors, useful in efficiently screening candidate
PT compounds for expression inducers of the genes, e.g. for treatment of
PT diabetes -
XX
PS Example 3; Page 30; 73pp; Japanese.
XX
CC The present invention describes an expression inducer (I) of hepatocyte
CC growth factor (HGF) gene and/or Reg gene containing a cytokine with
CC gp130 as receptor and glucocorticoid. Also described are: (1) a promoter
CC for pancreatic Langerhans beta cell proliferation containing a cytokine
CC with gp130 as receptor and glucocorticoid; (2) an apoptosis inhibitor
CC containing a cytokine with gp130 as receptor and glucocorticoid; (3) a
CC promoter for pancreatic Langerhans beta cells containing HGF or a nucleic
CC acid encoding HGF; (4) an apoptosis inhibitor of pancreatic Langerhans
CC beta cell proliferation containing HGF or a nucleic acid encoding HGF;
CC (5) drugs containing Reg protein or a nucleic acid encoding Reg protein;
CC (6) a method for screening candidate compounds as expression inducer of
CC Reg protein comprising: (a) supplying cells transferred with a vector
CC containing a reporter gene ligated functionally to the 5'-upstream of
CC Reg gene carrying poly(ADP ribose) synthase/polymerase (PARP) binding
CC sequence; (b) contacting the cells with a test sample and detecting

CC reporter activity of the cells; and (c) selecting a compound that can
CC increase the reporter activity as compared to a control; (7) a method
CC for screening candidate compounds as expression inducers of HGF gene
CC comprising: (a) supplying cells introduced with a vector containing a
CC reporter gene ligated functionally to 5' upstream of HGF gene carrying
CC IL-6/dexamethasone-responsive sequence; (b) contacting the cells with a
CC test sample and detecting reporter activity of the cells; and (c)
CC selecting a compound that can increase the reporter activity as compared
CC to a control. (I) has antidiabetic and apoptotic activities. The
CC promoters, inhibitors and screened compounds are for treating diabetes.
CC The present sequence represents a regulatory element related probe
CC which is used in an example from the present invention.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 24; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 13
ABZ58136
ID ABZ58136 standard; DNA; 22 BP.
XX
AC ABZ58136;
XX
DT 22-APR-2003 (first entry)
XX
DE Transcription factor Sp-1 probe.
XX
KW Human; osteocalcin; promoter; bone; tumour; prostate cancer;
KW metastasis; gene therapy; diagnosis; prognosis; marker; cytostatic;
KW transcription factor; Sp-1; probe; ss.
XX
OS Homo sapiens.
XX
PN WO2003006621-A2.
XX
PD 23-JAN-2003.
XX
PF 12-JUL-2002; 2002WO-US22216.
XX
PR 13-JUL-2001; 2001US-305360P.
XX
PA (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI Chung LWK, Yeung F;
XX
DR WPI; 2003-221733/21.
XX
PT Nucleic acid sequence for diagnosing, prognosing or treating calcified
PT tumors and tissues, e.g. prostate cancer, comprises an improved
PT recombinant human osteocalcin promoter activity -
XX
PS Example 1; Page 22; 55pp; English.
XX
CC The present sequence is that of a double-stranded probe for
CC transcription factor Sp-1, which was used in an electrophoretic
CC mobility shift assay in an example from the invention examining
CC prostate cancer-induced expression from the human osteocalcin
CC (hOC) promoter. The cis-acting element OSE-1 was shown to be a
CC weak Sp-1 binding site. An evaluation of the hOC promoter was
CC conducted in which the functional hierarchy of the elements OSE1,
CC OSE2 and AP-1/VDRE (vitamin D response element) was defined in
CC androgen-independent human prostate cancer cell line PC-3. By
CC juxtaposing dimers of these 3 elements, a minimal hOC super-promoter
CC (see ABZ58130) was produced, which displayed over 8-fold higher
CC activity than the native hOC promoter in a tissue-specific manner
CC in PC-3 cells. In one embodiment of the invention, the hOC
CC super-promoter is operably linked to a nucleic acid encoding a
CC heterologous protein, ribozyme, dominant-negative or antisense
CC RNA and used to deliver therapeutic genes to localised or
CC disseminated tumours. hOC promoters can also be used to deliver
CC therapeutic genes to fractured bones for bone repair. hOC promoter
CC activation by extracellular matrices and soluble factors secreted
CC by prostate cancer and bone cells in useful as a marker for the
CC diagnosis and prognosis of prostate cancer.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 25; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 14
AAF61841
ID AAF61841 standard; DNA; 49 BP.
XX
AC AAF61841;
XX
DT 26-JUL-2001 (first entry)
XX
DE SP1-specific DNA probe Zi-3.
XX
KW Probe; detection; electron donor; electron acceptor; competitive effect;
KW equilibrium constants determination; allele-specific interaction;
KW gene regulation; antigen identification; allosteric inhibitor; ss.
XX
OS Unidentified.
XX
PN WO200131057-A2.
XX
PD 03-MAY-2001.
XX
PF 17-OCT-2000; 2000WO-EP10209.
XX
PR 22-OCT-1999; 99DE-1050969.
XX
PA (AVET) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
XX
PI Muth J, Windhab N;
XX
DR WPI; 2001-335702/35.
XX
PT New nucleic acid probe bound to electrically conductive surface, used
PT e.g. to detect antigens or allosteric inhibitors, is attached to an
PT electron donor or acceptor -
XX
PS Example 6; Page 56; 57pp; German.
XX
CC This invention describes a novel nucleic acid probe (I) which comprises
CC a sequence, at least partly double-stranded (ds), bonded to a conductive
CC surface and having attached to it at least one electron donor (ED) or at
CC least one electron acceptor (EA). (I) are used to detect any type of
CC interaction, direct or indirect, that involves (I), also for
CC determination of equilibrium constants, competitive effects and rates of
CC reaction. Some typical applications are identification of antigens and
CC allosteric inhibitors; studying gene regulation (expression or
CC transcription), including allele-specific interactions between
CC regulatory factors and DNA sequences, also separation of interacting
CC factors. (I) are electrically 'readable' and provide rapid and simple
CC detection of any type of (in)direct interaction. Very small changes in
CC conduction can be detected. (I) have good compatibility with a variety
CC of reaction media, e.g. crude extracts can be analyzed. This sequence
CC represents a SP1-specific DNA probe, Zi-3 which is used to illustrate the
CC method of the invention.
XX

```
SQ   Sequence 49 BP; 6 A; 16 C; 16 G; 11 T; 0 other;

  Query Match             100.0%;  Score 22;  DB 22;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 15
AAF61841/c
ID   AAF61841 standard; DNA; 49 BP.
XX
AC   AAF61841;
XX
DT   26-JUL-2001  (first entry)
XX
DE   SP1-specific DNA probe Zi-3.
XX
KW   Probe; detection; electron donor; electron acceptor; competitive effect;
KW   equilibrium constants determination; allele-specific interaction;
KW   gene regulation; antigen identification; allosteric inhibitor; ss.
XX
OS   Unidentified.
XX
PN   WO200131057-A2.
XX
PD   03-MAY-2001.
XX
PF   17-OCT-2000; 2000WO-EP10209.
XX
PR   22-OCT-1999;   99DE-1050969.
XX
PA   (AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
XX
PI   Muth J,  Windhab N;
XX
DR   WPI; 2001-335702/35.
XX
PT   New nucleic acid probe bound to electrically conductive surface, used
PT   e.g. to detect antigens or allosteric inhibitors, is attached to an
PT   electron donor or acceptor  -
XX
PS   Example 6; Page 56; 57pp; German.
XX
CC   This invention describes a novel nucleic acid probe (I) which comprises
CC   a sequence, at least partly double-stranded (ds), bonded to a conductive
CC   surface and having attached to it at least one electron donor (ED) or at
CC   least one electron acceptor (EA). (I) are used to detect any type of
CC   interaction, direct or indirect, that involves (I), also for
CC   determination of equilibrium constants, competitive effects and rates of
CC   reaction. Some typical applications are identification of antigens and
CC   allosteric inhibitors; studying gene regulation (expression or
CC   transcription), including allele-specific interactions between
CC   regulatory factors and DNA sequences, also separation of interacting
CC   factors. (I) are electrically 'readable' and provide rapid and simple
CC   detection of any type of (in)direct interaction. Very small changes in
CC   conduction can be detected. (I) have good compatibility with a variety
CC   of reaction media, e.g. crude extracts can be analyzed. This sequence
CC   represents a SP1-specific DNA probe, Zi-3 which is used to illustrate the
CC   method of the invention.
XX
SQ   Sequence 49 BP; 6 A; 16 C; 16 G; 11 T; 0 other;

  Query Match             100.0%;  Score 22;  DB 22;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db         49 ATTCGATCGGGGCGGGGCGAGC 28


RESULT 16
AAF61843
ID   AAF61843 standard; DNA; 49 BP.
XX
AC   AAF61843;
XX
DT   26-JUL-2001  (first entry)
XX
DE   SP1-specific DNA probe Zi-5.
XX
KW   Probe; detection; electron donor; electron acceptor; competitive effect;
KW   equilibrium constants determination; allele-specific interaction;
KW   gene regulation; antigen identification; allosteric inhibitor; ss.
XX
OS   Unidentified.
XX
PN   WO200131057-A2.
XX
PD   03-MAY-2001.
XX
PF   17-OCT-2000; 2000WO-EP10209.
XX
PR   22-OCT-1999;   99DE-1050969.
XX
PA   (AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
XX
PI   Muth J,  Windhab N;
XX
DR   WPI; 2001-335702/35.
XX
PT   New nucleic acid probe bound to electrically conductive surface, used
PT   e.g. to detect antigens or allosteric inhibitors, is attached to an
PT   electron donor or acceptor  -
XX
PS   Example 6; Page 56; 57pp; German.
XX
CC   This invention describes a novel nucleic acid probe (I) which comprises
CC   a sequence, at least partly double-stranded (ds), bonded to a conductive
CC   surface and having attached to it at least one electron donor (ED) or at
CC   least one electron acceptor (EA). (I) are used to detect any type of
CC   interaction, direct or indirect, that involves (I), also for
CC   determination of equilibrium constants, competitive effects and rates of
CC   reaction. Some typical applications are identification of antigens and
CC   allosteric inhibitors; studying gene regulation (expression or
CC   transcription), including allele-specific interactions between
CC   regulatory factors and DNA sequences, also separation of interacting
CC   factors. (I) are electrically 'readable' and provide rapid and simple
CC   detection of any type of (in)direct interaction. Very small changes in
CC   conduction can be detected. (I) have good compatibility with a variety
CC   of reaction media, e.g. crude extracts can be analyzed. This sequence
CC   represents a SP1-specific DNA probe, Zi-5 which is used to illustrate the
CC   method of the invention.
XX
SQ   Sequence 49 BP; 6 A; 16 C; 16 G; 11 T; 0 other;

  Query Match             100.0%;  Score 22;  DB 22;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 17
AAF61843/c
ID   AAF61843 standard; DNA; 49 BP.
XX
AC   AAF61843;
XX
DT   26-JUL-2001  (first entry)
```

XX
DE SP1-specific DNA probe Zi-5.
XX
KW Probe; detection; electron donor; electron acceptor; competitive effect;
KW equilibrium constants determination; allele-specific interaction;
KW gene regulation; antigen identification; allosteric inhibitor; ss.
XX
OS Unidentified.
XX
PN WO200131057-A2.
XX
PD 03-MAY-2001.
XX
PF 17-OCT-2000; 2000WO-EP10209.
XX
PR 22-OCT-1999; 99DE-1050969.
XX
PA (AVET) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
XX
PI Muth J, Windhab N;
XX
DR WPI; 2001-335702/35.
XX
PT New nucleic acid probe bound to electrically conductive surface, used
PT e.g. to detect antigens or allosteric inhibitors, is attached to an
PT electron donor or acceptor -
XX
PS Example 6; Page 56; 57pp; German.
XX
CC This invention describes a novel nucleic acid probe (I) which comprises
CC a sequence, at least partly double-stranded (ds), bonded to a conductive
CC surface and having attached to it at least one electron donor (ED) or at
CC least one electron acceptor (EA). (I) are used to detect any type of
CC interaction, direct or indirect, that involves (I), also for
CC determination of equilibrium constants, competitive effects and rates of
CC reaction. Some typical applications are identification of antigens and
CC allosteric inhibitors; studying gene regulation (expression or
CC transcription), including allele-specific interactions between
CC regulatory factors and DNA sequences, also separation of interacting
CC factors. (I) are electrically 'readable' and provide rapid and simple
CC detection of any type of (in)direct interaction. Very small changes in
CC conduction can be detected. (I) have good compatibility with a variety
CC of reaction media, e.g. crude extracts can be analyzed. This sequence
CC represents a SP1-specific DNA probe, Zi-5 which is used to illustrate the
CC method of the invention.
XX
SQ Sequence 49 BP; 6 A; 16 C; 16 G; 11 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 49;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 49 ATTCGATCGGGGCGGGGCGAGC 28

RESULT 18
AAF99620
ID AAF99620 standard; DNA; 21 BP.
XX
AC AAF99620;
XX
DT 12-JUN-2001 (first entry)
XX
DE Immunostimulatory nucleic acid #736.
XX
KW Vaccine; cytostatic; virucidal; bactericidal; fungicidal; anti-parasitic;
KW immunostimulatory; tumour; viral infection; bacterial infection;
KW fungal infection; parasitic infection; cancer; asthma;
KW infectious disease; allergy; immune deficiency; phosphorothioate; ss.
XX
OS Synthetic.
XX
PN WO200122972-A2.
XX
PD 05-APR-2001.
XX
PF 25-SEP-2000; 2000WO-US26383.
XX
PR 25-SEP-1999; 99US-0156113.
PR 27-SEP-1999; 99US-0156135.
PR 23-AUG-2000; 2000US-0227436.
XX
PA (IOWA) UNIV IOWA RES FOUND.
PA (COLE-) COLEY PHARM GMBH.
XX
PI Krieg AM, Schetter C, Vollmer J;
XX
DR WPI; 2001-273485/28.
XX
PT Vaccinating against tumors, infectious diseases, allergies and asthma
PT using immunostimulatory Py-rich and TG nucleic acids -
XX
PS Claim 101; Page 54; 338pp; English.
XX
CC The present invention relates to a method for stimulating an immune
CC response. The method comprises administering an immunostimulatory nucleic
CC acid to a non-rodent subject in sufficient quantity to stimulate an
CC immune response. The present sequence is one such immunostimulatory
CC nucleic acid. The immunostimulatory nucleic acids can be pyrimidine rich
CC (py-rich) or thymidine (T) rich. The method is used to vaccinate subjects
CC against tumour antigens, viral antigens (e.g. herpesviridae, retroviridae
CC and/or orthomyxoviridae), bacterial antigens (e.g. toxoplasma,
CC haemophilus, campylobacter, clostridium, Escherichia coli and/or
CC staphylococcus), fungal antigens and/or parasitic antigens. The method is
CC also useful for preventing cancer, asthma, infectious disease, allergy or
CC immune deficiency. The present sequence can also be used to redirect a
CC Th2 to a Th1 immune response and to activate immune cells.
CC Note: the present sequence may have a phosphorothioate backbone.
XX
SQ Sequence 21 BP; 3 A; 4 C; 11 G; 3 T; 0 other;

Query Match 95.5%; Score 21; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
|||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAG 21

RESULT 19
AAF99621/c
ID AAF99621 standard; DNA; 21 BP.
XX
AC AAF99621;
XX
DT 12-JUN-2001 (first entry)
XX
DE Immunostimulatory nucleic acid #737.
XX
KW Vaccine; cytostatic; virucidal; bactericidal; fungicidal; anti-parasitic;
KW immunostimulatory; tumour; viral infection; bacterial infection;
KW fungal infection; parasitic infection; cancer; asthma;
KW infectious disease; allergy; immune deficiency; phosphorothioate; ss.
XX
OS Synthetic.
XX
PN WO200122972-A2.
XX
PD 05-APR-2001.
XX
PF 25-SEP-2000; 2000WO-US26383.
XX
PR 25-SEP-1999; 99US-0156113.

PR 27-SEP-1999; 99US-0156135.
PR 23-AUG-2000; 2000US-0227436.
XX
PA (IOWA) UNIV IOWA RES FOUND.
PA (COLE-) COLEY PHARM GMBH.
XX
PI Krieg AM, Schetter C, Vollmer J;
XX
DR WPI; 2001-273485/28.
XX
PT Vaccinating against tumors, infectious diseases, allergies and asthma
PT using immunostimulatory Py-rich and TG nucleic acids -
XX
PS Claim 101; Page 54; 338pp; English.
XX
CC The present invention relates to a method for stimulating an immune
CC response. The method comprises administering an immunostimulatory nucleic
CC acid to a non-rodent subject in sufficient quantity to stimulate an
CC immune response. The present sequence is one such immunostimulatory
CC nucleic acid. The immunostimulatory nucleic acids can be pyrimidine rich
CC (py-rich) or thymidine (T) rich. The method is used to vaccinate subjects
CC against tumour antigens, viral antigens (e.g. herpesviridae, retroviridae
CC and/or orthomyxoviridae), bacterial antigens (e.g. toxoplasma,
CC haemophilus, campylobacter, clostridium, Escherichia coli and/or
CC staphylococcus), fungal antigens and/or parasitic antigens. The method is
CC also useful for preventing cancer, asthma, infectious disease, allergy or
CC immune deficiency. The present sequence can also be used to redirect a
CC Th2 to a Th1 immune response and to activate immune cells.
CC Note: the present sequence may have a phosphorothioate backbone.
XX
SQ Sequence 21 BP; 3 A; 11 C; 4 G; 3 T; 0 other;

Query Match 95.5%; Score 21; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 21 ATTCGATCGGGGCGGGGCGAG 1

RESULT 20
ABS78341
ID ABS78341 standard; DNA; 21 BP.
XX
AC ABS78341;
XX
DT 13-DEC-2002 (first entry)
XX
DE Angiogenesis inhibitory oligonucleotide #825.
XX
KW Angiogenesis inhibitor; ss; angiogenesis; solid tumour growth;
KW tumour metastasis; precancerous lesion; rheumatoid arthritis;
KW psoriasis; diabetic retinopathy; retinopathy of prematurity;
KW macular degeneration; corneal graft rejection; neovascular glaucoma;
KW retrolental fibroplasia; rubeosis; Osler-Webber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophiliac joint; angiofibroma; wound granulation;
KW intestinal adhesion; atherosclerosis; scleroderma; hypertrophic scar.
XX
OS Synthetic.
XX
PN WO200253141-A2.
XX
PD 11-JUL-2002.
XX
PF 14-DEC-2001; 2001WO-US48458.
XX
PR 14-DEC-2000; 2000US-255534P.
XX
PA (COLE-) COLEY PHARM GROUP INC.
XX
PI Bratzler RL;
XX
DR WPI; 2002-566690/60.
XX
PT Inhibiting angiogenesis in a subject, involves administering at least
PT one antiangiogenic nucleic acid molecule to the subject -
XX
PS Claim 2; Page 34; 276pp; English.
XX
CC The invention relates to inhibiting angiogenesis in a subject, comprising
CC administering at least one antiangiogenic nucleic acid molecule.
CC Also included is a kit comprising a first container housing the
CC antiangiogenic nucleic acids, and instructions for administering them to
CC a subject having a condition characterised by unwanted angiogenesis.
CC The method is useful for inhibiting angiogenesis associated with solid
CC tumour growth, tumour metastasis, precancerous lesion, rheumatoid
CC arthritis, psoriasis, diabetic retinopathy, retinopathy of prematurity,
CC macular degeneration, corneal graft rejection, neovascular glaucoma,
CC retrolental fibroplasia, rubeosis, Osler-Webber Syndrome, myocardial
CC angiogenesis, plaque neovascularisation, telangiectasia, haemophiliac
CC joints, angiofibroma, wound granulation, intestinal adhesions,
CC atherosclerosis, scleroderma and hypertrophic scars. The present
CC sequence is an antiangiogenic nucleic acid of the invention.
XX
SQ Sequence 21 BP; 3 A; 4 C; 11 G; 3 T; 0 other;

Query Match 95.5%; Score 21; DB 24; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAG 21

RESULT 21
ABS78342/c
ID ABS78342 standard; DNA; 21 BP.
XX
AC ABS78342;
XX
DT 13-DEC-2002 (first entry)
XX
DE Angiogenesis inhibitory oligonucleotide #826.
XX
KW Angiogenesis inhibitor; ss; angiogenesis; solid tumour growth;
KW tumour metastasis; precancerous lesion; rheumatoid arthritis;
KW psoriasis; diabetic retinopathy; retinopathy of prematurity;
KW macular degeneration; corneal graft rejection; neovascular glaucoma;
KW retrolental fibroplasia; rubeosis; Osler-Webber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophiliac joint; angiofibroma; wound granulation;
KW intestinal adhesion; atherosclerosis; scleroderma; hypertrophic scar.
XX
OS Synthetic.
XX
PN WO200253141-A2.
XX
PD 11-JUL-2002.
XX
PF 14-DEC-2001; 2001WO-US48458.
XX
PR 14-DEC-2000; 2000US-255534P.
XX
PA (COLE-) COLEY PHARM GROUP INC.
XX
PI Bratzler RL;
XX
DR WPI; 2002-566690/60.
XX
PT Inhibiting angiogenesis in a subject, involves administering at least
PT one antiangiogenic nucleic acid molecule to the subject -
XX
PS Claim 2; Page 34; 276pp; English.

XX
CC The invention relates to inhibiting angiogenesis in a subject, comprising
CC administering at least one antiangiogenic nucleic acid molecule.
CC Also included is a kit comprising a first container housing the
CC antiangiogenic nucleic acids, and instructions for administering them to
CC a subject having a condition characterised by unwanted angiogenesis.
CC The method is useful for inhibiting angiogenesis associated with solid
CC tumour growth, tumour metastasis, precancerous lesion, rheumatoid
CC arthritis, psoriasis, diabetic retinopathy, retinopathy of prematurity,
CC macular degeneration, corneal graft rejection, neovascular glaucoma,
CC retrolental fibroplasia, rubeosis, Osler-Webber Syndrome, myocardial
CC angiogenesis, plaque neovascularisation, telangiectasia, haemophiliac
CC joints, angiofibroma, wound granulation, intestinal adhesions,
CC atherosclerosis, scleroderma and hypertrophic scars. The present
CC sequence is an antiangiogenic nucleic acid of the invention.
XX
SQ Sequence 21 BP; 3 A; 11 C; 4 G; 3 T; 0 other;

Query Match 95.5%; Score 21; DB 24; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 21 ATTCGATCGGGGCGGGGCGAG 1

RESULT 22
ABL38746
ID ABL38746 standard; DNA; 21 BP.
XX
AC ABL38746;
XX
DT 16-APR-2002 (first entry)
XX
DE Immunostimulatory nucleic acid SEQ ID NO: 115.
XX
KW Antibody-induced cell lysis; cancer; immunostimulatory; CD20;
KW angiogenesis; metastasis; cytostatic; ss.
XX
OS Synthetic.
XX
PN WO200197843-A2.
XX
PD 27-DEC-2001.
XX
PF 22-JUN-2001; 2001WO-US20154.
XX
PR 22-JUN-2000; 2000US-213346P.
XX
PA (IOWA) UNIV IOWA RES FOUND.
XX
PI Weiner G, Hartmann G;
XX
DR WPI; 2002-154611/20.
XX
PT Treating or preventing cancer, such as basal cell carcinoma, comprises
PT administering immunostimulatory nucleic acids that induce expression of
PT cell surface antigens and antibodies to a subject having or at risk of
PT developing cancer -
XX
PS Disclosure; Page 124; 312pp; English.
XX
CC The present invention relates to methods for treating or preventing
CC cancer, involving administering to a subject having or at risk of
CC developing cancer immunostimulatory nucleic acids that induce expression
CC of cell surface antigens and antibodies. The methods are useful for
CC treating or preventing cancer such as basal cell carcinoma, bladder
CC cancer, bone cancer, brain and central nervous system (CNS) cancer,
CC breast cancer, cervical cancer, colon and rectum cancer, connective
CC tissue cancer, oesophageal cancer, eye cancer, kidney cancer, larynx
CC cancer, leukaemia, liver cancer, lung cancer, Hodgkin's lymphoma,
CC non-Hodgkin's lymphoma, melanoma, myeloma, oral cavity cancer, ovarian
CC cancer, pancreatic cancer, prostate cancer, rhabdomyosarcoma, skin
CC cancer, stomach cancer, testicular cancer, and uterine cancer. The
CC present sequence is an immunostimulatory oligonucleotide described in
CC the exemplification of the invention.
XX
SQ Sequence 21 BP; 3 A; 4 C; 11 G; 3 T; 0 other;

Query Match 95.5%; Score 21; DB 24; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAG 21

RESULT 23
ABL38839/c
ID ABL38839 standard; DNA; 21 BP.
XX
AC ABL38839;
XX
DT 16-APR-2002 (first entry)
XX
DE Immunostimulatory nucleic acid SEQ ID NO: 229.
XX
KW Antibody-induced cell lysis; cancer; immunostimulatory; CD20;
KW angiogenesis; metastasis; cytostatic; ss.
XX
OS Synthetic.
XX
PN WO200197843-A2.
XX
PD 27-DEC-2001.
XX
PF 22-JUN-2001; 2001WO-US20154.
XX
PR 22-JUN-2000; 2000US-213346P.
XX
PA (IOWA) UNIV IOWA RES FOUND.
XX
PI Weiner G, Hartmann G;
XX
DR WPI; 2002-154611/20.
XX
PT Treating or preventing cancer, such as basal cell carcinoma, comprises
PT administering immunostimulatory nucleic acids that induce expression of
PT cell surface antigens and antibodies to a subject having or at risk of
PT developing cancer -
XX
PS Disclosure; Page 153; 312pp; English.
XX
CC The present invention relates to methods for treating or preventing
CC cancer, involving administering to a subject having or at risk of
CC developing cancer immunostimulatory nucleic acids that induce expression
CC of cell surface antigens and antibodies. The methods are useful for
CC treating or preventing cancer such as basal cell carcinoma, bladder
CC cancer, bone cancer, brain and central nervous system (CNS) cancer,
CC breast cancer, cervical cancer, colon and rectum cancer, connective
CC tissue cancer, oesophageal cancer, eye cancer, kidney cancer, larynx
CC cancer, leukaemia, liver cancer, lung cancer, Hodgkin's lymphoma,
CC non-Hodgkin's lymphoma, melanoma, myeloma, oral cavity cancer, ovarian
CC cancer, pancreatic cancer, prostate cancer, rhabdomyosarcoma, skin
CC cancer, stomach cancer, testicular cancer, and uterine cancer. The
CC present sequence is an immunostimulatory oligonucleotide described in
CC the exemplification of the invention.
XX
SQ Sequence 21 BP; 3 A; 11 C; 4 G; 3 T; 0 other;

Query Match 95.5%; Score 21; DB 24; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db         21 ATTCGATCGGGGCGGGGCGAG 1


RESULT 24
ABT17254
ID   ABT17254 standard; DNA; 21 BP.
XX
AC   ABT17254;
XX
DT   10-APR-2003  (first entry)
XX
DE   Transcription factor probe - SEQ ID No 81.
XX
KW   Probe; ss; transcription factor-protein complex; transcription factor;
KW   drug screening; drug identification; array hybridisation.
XX
OS   Unidentified.
XX
PN   WO2002101351-A2.
XX
PD   19-DEC-2002.
XX
PF   30-MAY-2002; 2002WO-US17408.
XX
PR   08-JUN-2001; 2001US-0877243.
PR   08-JUN-2001; 2001US-0877403.
PR   08-JUN-2001; 2001US-0877705.
PR   08-JUN-2001; 2001US-0877738.
PR   05-SEP-2001; 2001US-0947274.
XX
PA   (PANO-) PANOMICS INC.
XX
PI   Li X;
XX
DR   WPI; 2003-148829/14.
XX
PT   Identifying transcription factor-protein complexes, by isolating
PT   transcription factor complexes from sample based on a specific type of
PT   factor, and identifying different proteins present in isolated
PT   complexes  -
XX
PS   Disclosure; Fig 6; 167pp; English.
XX
CC   The invention comprises a method for identifying complexes between a
CC   transcription factor and another protein. The invention also comprises a
CC   method for isolating DNA probes which bind to activated transcription
CC   factors. The methods of the invention are useful for identifying
CC   transcription factor-protein interactions. The methods of the invention
CC   are also useful for facilitating the screening and identification of new
CC   drugs, characterising their mechanism of action and screening for adverse
CC   side effects based on drug's impact expression. The present DNA sequence
CC   represents a probe used in the method of the invention.
XX
SQ   Sequence 21 BP; 3 A; 4 C; 11 G; 3 T; 0 other;

  Query Match             95.5%;  Score 21;  DB 25;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 8.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAG 21


RESULT 25
ABT17255/c
ID   ABT17255 standard; DNA; 21 BP.
XX
AC   ABT17255;
XX
DT   10-APR-2003  (first entry)
XX
DE   Transcription factor probe - SEQ ID No 82.
XX
KW   Probe; ss; transcription factor-protein complex; transcription factor;
KW   drug screening; drug identification; array hybridisation.
XX
OS   Unidentified.
XX
PN   WO2002101351-A2.
XX
PD   19-DEC-2002.
XX
PF   30-MAY-2002; 2002WO-US17408.
XX
PR   08-JUN-2001; 2001US-0877243.
PR   08-JUN-2001; 2001US-0877403.
PR   08-JUN-2001; 2001US-0877705.
PR   08-JUN-2001; 2001US-0877738.
PR   05-SEP-2001; 2001US-0947274.
XX
PA   (PANO-) PANOMICS INC.
XX
PI   Li X;
XX
DR   WPI; 2003-148829/14.
XX
PT   Identifying transcription factor-protein complexes, by isolating
PT   transcription factor complexes from sample based on a specific type of
PT   factor, and identifying different proteins present in isolated
PT   complexes  -
XX
PS   Disclosure; Fig 6; 167pp; English.
XX
CC   The invention comprises a method for identifying complexes between a
CC   transcription factor and another protein. The invention also comprises a
CC   method for isolating DNA probes which bind to activated transcription
CC   factors. The methods of the invention are useful for identifying
CC   transcription factor-protein interactions. The methods of the invention
CC   are also useful for facilitating the screening and identification of new
CC   drugs, characterising their mechanism of action and screening for adverse
CC   side effects based on drug's impact expression. The present DNA sequence
CC   represents a probe used in the method of the invention.
XX
SQ   Sequence 21 BP; 3 A; 11 C; 4 G; 3 T; 0 other;

  Query Match             95.5%;  Score 21;  DB 25;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 8.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db         21 ATTCGATCGGGGCGGGGCGAG 1


RESULT 26
ABT17322/c
ID   ABT17322 standard; DNA; 63 BP.
XX
AC   ABT17322;
XX
DT   10-APR-2003  (first entry)
XX
DE   Transcription factor-related array hybridisation probe - SEQ ID No 149.
XX
KW   Probe; ss; transcription factor-protein complex; transcription factor;
KW   drug screening; drug identification; array hybridisation.
XX
OS   Unidentified.
XX
PN   WO2002101351-A2.
XX
PD   19-DEC-2002.
XX
```

PF 30-MAY-2002; 2002WO-US17408.
XX
PR 08-JUN-2001; 2001US-0877243.
PR 08-JUN-2001; 2001US-0877403.
PR 08-JUN-2001; 2001US-0877705.
PR 08-JUN-2001; 2001US-0877738.
PR 05-SEP-2001; 2001US-0947274.
XX
PA (PANO-) PANOMICS INC.
XX
PI Li X;
XX
DR WPI; 2003-148829/14.
XX
PT Identifying transcription factor-protein complexes, by isolating
PT transcription factor complexes from sample based on a specific type of
PT factor, and identifying different proteins present in isolated
PT complexes -
XX
PS Disclosure; Fig 6; 167pp; English.
XX
CC The invention comprises a method for identifying complexes between a
CC transcription factor and another protein. The invention also comprises a
CC method for isolating DNA probes which bind to activated transcription
CC factors. The methods of the invention are useful for identifying
CC transcription factor-protein interactions. The methods of the invention
CC are also useful for facilitating the screening and identification of new
CC drugs, characterising their mechanism of action and screening for adverse
CC side effects based on drug's impact expression. The present DNA sequence
CC represents a probe used in the method of the invention.
XX
SQ Sequence 63 BP; 9 A; 33 C; 12 G; 9 T; 0 other;

Query Match 95.5%; Score 21; DB 25; Length 63;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 63 ATTCGATCGGGGCGGGGCGAG 43

RESULT 27
ABK89728
ID ABK89728 standard; DNA; 22 BP.
XX
AC ABK89728;
XX
DT 05-NOV-2002 (first entry)
XX
DE Oestrogen response element #11.
XX
KW Oestrogen receptor alpha; breast cancer; pre-malignant lesion;
KW invasive breast cancer; A908G oestrogen receptor alpha transition;
KW oestrogen receptor alpha K303R substitution;
KW oestrogen response element; ds.
XX
OS Synthetic.
XX
PN WO200257283-A1.
XX
PD 25-JUL-2002.
XX
PF 16-JAN-2002; 2002WO-US04982.
XX
PR 19-JAN-2001; 2001US-262990P.
PR 09-JUL-2001; 2001US-304018P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Fuqua S, O'Connell P, Allred DC, Hopp TA;
XX
DR WPI; 2002-590711/63.
XX
PT New isolated oestrogen receptor alpha with A908G mutation or K303R
PT substitution, useful as diagnostic marker in breast tissue such as
PT pre-malignant lesions for the development of breast cancer,
PT particularly invasive breast cancer -
XX
PS Claim 43; Page 90; 133pp; English.
XX
CC The invention relates to an isolated oestrogen receptor alpha nucleic
CC acid sequence comprising an A908G mutation, or an amino acid sequence
CC comprising a K303R substitution. Also described are methods for
CC detecting susceptibility to development of breast cancer or invasive
CC breast cancer in an individual, for diagnosing breast cancer in an
CC individual; and for screening for a modulator of an oestrogen receptor
CC alpha polypeptide comprising a K303R substitution. The oestrogen receptor
CC alpha is useful as a diagnostic marker in breast tissue such as pre-
CC malignant lesions for the development of breast cancer, particularly
CC invasive breast cancer. The methods are useful for determining
CC susceptibility to development of breast cancer, for diagnosing,
CC preventing or treating breast cancer. Transgenic mice may be used for
CC screening and identifying agents that interact with the oestrogen
CC receptor alpha, or affect breast tissue health. The A908G oestrogen
CC receptor alpha transition is frequently present in pre-malignant lesions
CC of the breast and can occur in the adjacent normal-appearing breast
CC epithelium. The present sequence represents an oestrogen response
CC element used in a reporter vector to screen for antagonists and agonists
CC of the oestrogen receptor alpha K303R polypeptide.
XX
SQ Sequence 22 BP; 4 A; 5 C; 10 G; 3 T; 0 other;

Query Match 92.7%; Score 20.4; DB 24; Length 22;
Best Local Similarity 95.5%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||| ||||||||||||
Db 1 ATTCGATCAGGGCGGGGCGAGC 22

RESULT 28
AAV46004
ID AAV46004 standard; DNA; 20 BP.
XX
AC AAV46004;
XX
DT 16-OCT-1998 (first entry)
XX
DE Immune adjuvant SP-1.
XX
KW Immune system; adjuvant; vaccine; cancer; prophylactic; pathogenicity;
KW modulator; tolerance; regulator; helper cell; antigen; immunoglobulin;
KW Ig class; autoimmune response; T-cell; B-cell; tumour; ss.
XX
OS Class Bacteria.
XX
PN EP855184-A1.
XX
PD 29-JUL-1998.
XX
PF 23-JAN-1997; 97EP-0101019.
XX
PR 23-JAN-1997; 97EP-0101019.
XX
PA (HEEG/) HEEG K.
PA (LIPF/) LIPFORD G B.
PA (WAGN/) WAGNER H.
XX
PI Heeg K, Lipford GB, Wagner H;
XX
DR WPI; 1998-389630/34.
XX
PT Antigenic composition comprises polynucleotide fragment and antigen
PT - used as vaccine to treat or prevent e.g. cancer or pathogen

PT infections and to modulate immune response e.g. tolerance break and
PT regulation of TH1/TH2 cells
XX
PS Example 5; Page 9; 28pp; English.
XX
CC AAV45993-V46019 are fragments of bacterial polynucleotides which are
CC used as immune adjuvants for inclusion into vaccines to treat cancer and
CC for prophylaxis and/or treatment of conditions caused by pathogenic
CC micro-organisms. The polynucleotide is used for modulation of an immune
CC response and the modulation is selected from the group break of
CC tolerance, regulation of TH1/TH2 helper cell responses, switch of Ig
CC classes, treatment of autoimmune responses and induction of tolerances.
CC DNA oligomers are used to enhance the reactivity of immune cells to
CC viral, bacterial and parasitic antigens, to break tolerance in anergic T
CC and B cells e.g. against tumour antigens, as adjuvants in vaccination
CC against tumour-defined antigens and immunostimulatory substances in an
CC immune response against tumours and to suppress immune reactions of the
CC innate and acquired immune system. The composition is inexpensive and
CC stable and does not cause lethal shock, which happens with prior art
CC bacterial sequences.
XX
SQ Sequence 20 BP; 2 A; 5 C; 11 G; 2 T; 0 other;

Query Match 90.9%; Score 20; DB 19; Length 20;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||
Db 1 TCGATCGGGGCGGGGCGAGC 20

RESULT 29
AAZ99627
ID AAZ99627 standard; DNA; 20 BP.
XX
AC AAZ99627;
XX
DT 12-JUL-2000 (first entry)
XX
DE Nucleotide sequence of G-motif oligonucleotide SP1.
XX
KW G-motif oligonucleotide; vaccine; Toxoplasmosis; viral infection;
KW antigen presenting cell activation; natural killer cell; septic shock;
KW cytotoxic T-lymphocyte; inflammation; autoimmune disease;
KW rheumatoid arthritis; Crohn's disease; sarcoidosis; multiple sclerosis;
KW Kawasaki syndrome; graft-versus-host disease; transplant rejection;
KW helper T cell response 1-mediated disease; Lyme arthritis;
KW Streptococcal induced arthritis; chronic inflammatory bowel disease;
KW psoriasis vulgaris; experimental allergic encephalomyelitis;
KW insulin-dependent diabetes mellitus; bacterial infection;
KW parasitic infection; Leishmaniasis; spontaneous abortion; tumour; ss.
XX
OS Synthetic.
XX
PN WO200014217-A2.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-EP06502.
XX
PR 03-SEP-1998; 98EP-0116652.
XX
PA (CPGI-) CPG IMMUNOPHARMACEUTICALS GMBH.
XX
PI Wagner H, Lipford GB, Heeg K;
XX
DR WPI; 2000-256970/22.
XX
PT Compositions comprising G-motif oligonucleotides useful for treating
PT e.g. septic shock, rheumatoid arthritis, diabetes and human
PT immunodeficiency virus infections -
XX
PS Example 14; Page 32; 75pp; English.
XX
CC The present sequence represents a G-motif oligonucleotide of the
CC invention. The specification describes compositions comprising G-motif
CC oligonucleotides. The G-motif oligonucleotides inhibit activation of
CC antigen presenting cells by inhibiting the uptake of DNA by a cell, by
CC stimulating natural killer cells, or by co-stimulating cytotoxic
CC T-lymphocytes. The G-motif oligonucleotides may be used for the
CC productions of vaccines for treating septic shock, inflammation,
CC autoimmune diseases (e.g. rheumatoid arthritis, Crohn's disease,
CC sarcoidosis, multiple sclerosis, Kawasaki syndrome, graft-versus-host
CC disease and transplant rejection), helper T cell response 1-mediated
CC diseases (e.g. Streptococcal induced arthritis, Lyme arthritis, chronic
CC inflammatory bowel disease, psoriasis vulgaris, experimental allergic
CC encephalomyelitis, and insulin-dependent diabetes mellitus), bacterial
CC infections, parasitic infections (e.g. Leishmaniasis or Toxoplasmosis),
CC viral infections (e.g. Cytomegalovirus and human immunodeficiency virus
CC (HIV)-infections), spontaneous abortions and tumours. They may also be
CC used to induce proliferation of bone marrow cells, especially macrophage
CC precursor cells.
XX
SQ Sequence 20 BP; 2 A; 5 C; 11 G; 2 T; 0 other;

Query Match 90.9%; Score 20; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||
Db 1 TCGATCGGGGCGGGGCGAGC 20

RESULT 30
AAL39229
ID AAL39229 standard; DNA; 20 BP.
XX
AC AAL39229;
XX
DT 05-SEP-2002 (first entry)
XX
DE Murine Toll-like receptor related CpG DNA SEQ ID No 104.
XX
KW Murine Toll-like receptor; TLR9; TLR7; TLR8; ISNA; ds.
XX
OS Unidentified.
XX
PN WO200222809-A2.
XX
PD 21-MAR-2002.
XX
PF 17-SEP-2001; 2001WO-US29229.
XX
PR 15-SEP-2000; 2000US-233035P.
PR 23-JAN-2001; 2001US-263657P.
PR 17-MAY-2001; 2001US-291726P.
PR 22-JUN-2001; 2001US-300210P.
XX
PA (COLE-) COLEY PHARM GMBH.
XX
PI Bauer S, Lipford G, Wagner H;
XX
DR WPI; 2002-393964/42.
XX
PT New isolated murine Toll-like receptor (TLR)9, TLR7, TLR8 polypeptides,
PT useful for identifying species specificity of immunostimulatory nucleic
PT acid and identifying immunostimulatory nucleic acids -
XX
PS Disclosure; Page 77; 195pp; English.
XX
CC The invention relates to isolated murine Toll-like receptors (TLR)9,
CC TLR7 and TLR8 polypeptides. These polypeptides comprise fully defined
CC sequences of 1032, 1050 or 1032 amino acids as given in specification, or
CC their fragments, where TLR9, TLR7 and TLR8 polypeptides or their

CC fragments have an amino acid sequence which is identical to human TLR9,
CC TLR7 or TLR8 polypeptides or their fragment except for at least one amino
CC acid of a murine TLR polypeptide. The isolated nucleic acids of the
CC invention are useful for inhibiting TLR9 signalling activity in a cell.
CC TLR7, TLR8 and TLR9 polypeptides are useful for identifying nucleic acid
CC molecules which interact with a TLR polypeptide or its fragment. The
CC TLR7, TLR8 or TLR9 polypeptides are also useful for identifying ISNA. The
CC TLR7, TLR8 and TLR9 polypeptides are also useful for comparing TLR9
CC signalling activity of a test compound (that is not a nucleic acid, and
CC is a polypeptide or a part of a combinatorial library of compounds) with
CC an ISNA. The TLR7, TLR8 and TLR9 polypeptides are also useful for
CC identifying species specificity of an ISNA. The isolated nucleic acids of
CC the invention are useful as probes or primers. This polynucleotide
CC sequence represents DNA relating to the isolated Toll-like receptors of
CC the invention.
XX
SQ Sequence 20 BP; 2 A; 5 C; 11 G; 2 T; 0 other;

Query Match 90.9%; Score 20; DB 24; Length 20;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||
Db 1 TCGATCGGGGCGGGGCGAGC 20

RESULT 31
ABA92272
ID ABA92272 standard; DNA; 22 BP.
XX
AC ABA92272;
XX
DT 10-JUN-2002 (first entry)
XX
DE Sp-1 mutant oligonucleotide, used in EMSA.
XX
KW SP-1; neuron; antialzheimers; antiparkinsonian; antisclerotic;
KW neuroprotective; nootropic; anticonvulsant; vascular; hypotensive;
KW cerebroprotective; virucide; anti-HIV; diagnosis; therapy;
KW electrophoretic mobility shift assay; EMSA; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200215912-A1.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26527.
XX
PR 25-AUG-2000; 2000US-228201P.
PR 26-OCT-2000; 2000US-243295P.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Ratan RR, Chatterjee S;
XX
DR WPI; 2002-242023/29.
XX
PT Diagnosing and treating diseases associated with oxidative stress, DNA
PT damage or growth factor depletion, e.g. Alzheimer's and Parkinson's, by
PT administering e.g. mithramycin, chromomycin, daunomycin, olivomycin and
PT WP631 -
XX
PS Example 1; Page 23; 69pp; English.
XX
CC The present sequence is that of a double-stranded mutated Sp-1
CC oligonucleotide, which was used in an electrophoretic mobility
CC shift assay to determine the effect of oxidative stress on Sp-1 DNA
CC binding, and to determine the effects of candidate compounds on Sp-1
CC protein levels. SP-1 DNA binding activity in cortical neurons was
CC shown to be low, but was dramatically enhanced by oxidative stress.
CC The invention provides methods for detecting and treating diseases
CC associated with oxidative stress, DNA damage or growth factor
CC depletion, and identifying agents for the treatment of such
CC conditions. A compound is deemed to be an inhibitor of oxidative
CC stress, DNA damage, growth factor depletion or cell death if it
CC reduces the protein level of an Sp family member or if it decreases
CC the binding of an Sp family member to DNA. A method for preventing
CC or treating a disease or disorder of the nervous system, the ageing
CC process or associated with apoptosis involves administering a
CC compound that inhibits the induction of an Sp family member or the
CC binding of an Sp family member to DNA, e.g. mithramycin,
CC chromomycin, daunomycin, olivomycin or WP631. Diseases and
CC disorders that can be treated include Alzheimer's disease,
CC Creutzfeldt-Jacob disease, kuru, Huntington's disease, aneurysm,
CC stroke associated with an increase in blood pressure, spinal cord
CC disease, spinal cord injury, brain injury, multiple system atrophy,
CC amylotrophic lateral sclerosis, progressive supranuclear palsy,
CC neurodegeneration associated with the ageing process, mitochondrial
CC disease, HIV infection, herpes infection and multiple sclerosis
CC (all claimed).
XX
SQ Sequence 22 BP; 3 A; 5 C; 9 G; 5 T; 0 other;

Query Match 85.5%; Score 18.8; DB 24; Length 22;
Best Local Similarity 90.9%; Pred. No. 71;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
|||||||||| ||||||||||
Db 1 ATTCGATCGGTTCGGGGCGAGC 22

RESULT 32
AAQ67304
ID AAQ67304 standard; DNA; 22 BP.
XX
AC AAQ67304;
XX
DT 30-MAR-1995 (first entry)
XX
DE Detection probe for DNA binding activity of Sp1 protein.
XX
KW Jun protein; Fos protein; Sp1 protein; DNA binding protein;
KW probe; detection; calibration; ss.
XX
OS Synthetic.
XX
PN JP06201692-A.
XX
PD 22-JUL-1994.
XX
PF 22-SEP-1993; 93JP-0257482.
XX
PR 22-SEP-1992; 92JP-0278126.
XX
PA (YAWH) NIPPON STEEL CHEM CO.
PA (YAWA) NIPPON STEEL CORP.
XX
DR WPI; 1994-273729/34.
XX
PT Probe for detection or calibration of DNA binding - and detection
PT or calibration of the protein by using it in situ
XX
PS Claim 5; Page 5; 6pp; Japanese.
XX
CC The probes given in AAQ67303-04 are used for the detection or
CC calibration of DNA binding and contain nucleotide sequences binding
CC specifically with DNA binding protein (Jun or Fos protein, and Sp1
CC protein respectively). Distribution of DNA binding protein in
CC tissues or cells can be inspected in situ, by using DNA binding
CC activity as indicator.
XX
SQ Sequence 22 BP; 3 A; 5 C; 11 G; 3 T; 0 other;

```
  Query Match            83.6%;  Score 18.4;  DB 15;  Length 22;
  Best Local Similarity   95.0%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              |||||||||||||||||| |
Db          3 TCGATCGGGGCGGGGCGATC 22


RESULT 33
AAQ30483
ID   AAQ30483 standard; DNA; 45 BP.
XX
AC   AAQ30483;
XX
DT   25-MAR-2003  (updated)
DT   19-MAR-1993  (first entry)
XX
DE   Oligonucleotide contg. Sp1 recognition sequence.
XX
KW   Control recognition element; decoy; cellular RNA; dumbbell;
KW   promoter; hormone receptor element; viral; liver; tissue; viral;
KW   proliferation; linker; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_structure  1..45
FT                   /*tag=  a
FT                   /note= "looped self structure"
XX
PN   WO9218522-A1.
XX
PD   29-OCT-1992.
XX
PF   17-APR-1992;   92WO-US03205.
XX
PR   18-APR-1991;   91US-0687337.
XX
PA   (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
PI   Chu BC,  Orgel L;
XX
DR   WPI; 1992-382035/46.
XX
PT   New oligo-nucleotide(s) contg. transcription control recognition
PT   element - stabilised by covalent bonding of two DNA strands, act
PT   as decoys for regulatory protein to modulate specific RNA
XX
PS   Example 3; Page 27; 41pp; English.
XX
CC   The oligonucleotide contains the double stranded control recognition
CC   element Sp1 recognition sequence, and was sythesised in this linear
CC   form for prodn. of dumbbell oligonucleotides.  The oligomer forms
CC   a looped self structure with a nicked gap between the 3'-OH and 5'-OH
CC   tail ends.  Certian bases may optionally have incorporated a
CC   phosphorothioate diester linkage instead of the normal phosphodiester
CC   bond.  The oligonucleotide acts as a decoy for proteins which
CC   exert transcription control, and so modulate specific cellular RNAs.
CC   Typical CREs which may be regulated include promoters, hormone
CC   receptor elements, viral, cellular, liver or tissue elements, etc, and
CC   a typical application is inhibition of viral proliferation.
CC   See also AAQ30472-518.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 45 BP; 3 A; 21 C; 14 G; 7 T; 0 other;

  Query Match            81.8%;  Score 18;  DB 13;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGA 20
                ||||||||||||||||||
Db         28 TCGATCGGGGCGGGGCGA 45


RESULT 34
AAQ30483/c
ID   AAQ30483 standard; DNA; 45 BP.
XX
AC   AAQ30483;
XX
DT   25-MAR-2003  (updated)
DT   19-MAR-1993  (first entry)
XX
DE   Oligonucleotide contg. Sp1 recognition sequence.
XX
KW   Control recognition element; decoy; cellular RNA; dumbbell;
KW   promoter; hormone receptor element; viral; liver; tissue; viral;
KW   proliferation; linker; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_structure  1..45
FT                   /*tag=  a
FT                   /note= "looped self structure"
XX
PN   WO9218522-A1.
XX
PD   29-OCT-1992.
XX
PF   17-APR-1992;   92WO-US03205.
XX
PR   18-APR-1991;   91US-0687337.
XX
PA   (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
PI   Chu BC,  Orgel L;
XX
DR   WPI; 1992-382035/46.
XX
PT   New oligo-nucleotide(s) contg. transcription control recognition
PT   element - stabilised by covalent bonding of two DNA strands, act
PT   as decoys for regulatory protein to modulate specific RNA
XX
PS   Example 3; Page 27; 41pp; English.
XX
CC   The oligonucleotide contains the double stranded control recognition
CC   element Sp1 recognition sequence, and was sythesised in this linear
CC   form for prodn. of dumbbell oligonucleotides.  The oligomer forms
CC   a looped self structure with a nicked gap between the 3'-OH and 5'-OH
CC   tail ends.  Certian bases may optionally have incorporated a
CC   phosphorothioate diester linkage instead of the normal phosphodiester
CC   bond.  The oligonucleotide acts as a decoy for proteins which
CC   exert transcription control, and so modulate specific cellular RNAs.
CC   Typical CREs which may be regulated include promoters, hormone
CC   receptor elements, viral, cellular, liver or tissue elements, etc, and
CC   a typical application is inhibition of viral proliferation.
CC   See also AAQ30472-518.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 45 BP; 3 A; 21 C; 14 G; 7 T; 0 other;

  Query Match            81.8%;  Score 18;  DB 13;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||
Db         24 GATCGGGGCGGGGCGAGC 7


RESULT 35
AAQ30484/c
```

```
ID   AAQ30484 standard; DNA; 47 BP.
XX
AC   AAQ30484;
XX
DT   25-MAR-2003  (updated)
DT   19-MAR-1993  (first entry)
XX
DE   Oligonucleotide contg. Sp1 recognition sequence.
XX
KW   Control recognition element; decoy; cellular RNA; dumbbell;
KW   promoter; hormone receptor element; viral; liver; tissue; viral;
KW   proliferation; linker; cyclic; closed circular.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_structure  1..47
FT                   /*tag=  a
FT                   /note= "closed, circular, self complementary
FT                          structure"
XX
PN   WO9218522-A1.
XX
PD   29-OCT-1992.
XX
PF   17-APR-1992;   92WO-US03205.
XX
PR   18-APR-1991;   91US-0687337.
XX
PA   (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
PI   Chu BC,  Orgel L;
XX
DR   WPI; 1992-382035/46.
XX
PT   New oligo-nucleotide(s) contg. transcription control recognition
PT   element - stabilised by covalent bonding of two DNA strands, act
PT   as decoys for regulatory protein to modulate specific RNA
XX
PS   Example 3; Page 27; 41pp; English.
XX
CC   The oligonucleotide contains the double stranded control recognition
CC   element Sp1 recognition sequence in the form of a dumbbell structure.
CC   The oligomer, when synthesised in its linear form forms a looped
CC   self structure with a nicked gap between the 3'-OH and 5'-OH tail
CC   ends, which may be phosphorylated with T4 polynucleotide kinase and
CC   ligated with DNA ligase to give a closed circular self complementary
CC   structure. The oligonucleotide acts as a decoy for proteins which
CC   exert transcription control, and so modulate specific cellular RNAs.
CC   Typical CREs which may be regulated include promoters, hormone
CC   receptor elements, viral, cellular, liver or tissue elements, etc, and
CC   a typical application is inhibition of viral proliferation.
CC   See also AAQ30472-518.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 47 BP; 3 A; 21 C; 16 G; 7 T; 0 other;

  Query Match             81.8%;  Score 18;  DB 13;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||
Db         42 GATCGGGGCGGGGCGAGC 25


RESULT 36
ABA00746/c
ID   ABA00746 standard; DNA; 183 BP.
XX
AC   ABA00746;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Enh1 enhancer.
XX
KW   Enhancer; Enh1; lentivirus; producer cell; gag; pol; tissue-specific;
KW   transfer vector; blood clotting factor; tat; bleeding disorder; env;
KW   hepatocyte transcription factor; HNF1; HNF3; HNF4; HNF6; haemophilia;
KW   liver-specific; transgene delivery; gene therapy; EBP; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        11..161
FT                   /*tag=  a
FT                   /note= "Hepatocyte derived elements"
XX
PN   WO200292134-A1.
XX
PD   21-NOV-2002.
XX
PF   14-MAY-2002; 2002WO-US15062.
XX
PR   14-MAY-2001; 2001US-291083P.
XX
PA   (CELL-) CELL GENESYS INC.
XX
PI   McArthur JG,  Talbot DJ,  Simmons AD,  McGuinness R,  Kelly M;
PI   Tsui LV,  Dull T;
XX
DR   WPI; 2003-120615/11.
XX
PT   New lentiviral producer cell, useful for optimizing liver-specific
PT   transgene delivery for treating bleeding disorders, e.g. hemophilia, or
PT   for tissue-specific gene therapy  -
XX
PS   Claim 11; Page 39; 61pp; English.
XX
CC   This sequence represents a cloned enhancer element (Enh1) which was
CC   used in the production of the lentiviral producer cell of the invention.
CC   The producer cell comprises:
CC   (a) a first nucleotide sequence comprising a gag, a pol, or gag and pol
CC   genes;
CC   (b) a second nucleotide sequence comprising a heterologous env gene;
CC   and
CC   (c) a third nucleotide sequence comprising a lentiviral transfer vector
CC   that comprises a gene that encodes a blood clotting factor operably
CC   linked to an expression control sequence. The producer cell lacks a
CC   functional tat gene.  This enhancer sequence was created by ligating
CC   together five synthetic oligonucleotide binding sites for hepatocyte
CC   transcription factors.  This enhancer element augmented expression
CC   levels up to two orders of magnitude over the enhancer-less vector. Enh1
CC   comprises the following elements:
CC   HNF1(sense) - HNF3(sense) - HNF4(antisense) - HNF1(antisense) -
CC   HNF6(sense) - EBP(antisense) - HNF4(antisense).
CC   The lentiviral producer cell and transfer vectors are useful for
CC   optimizing liver-specific transgene delivery for treating bleeding
CC   disorders, e.g. haemophilia. The vectors are also useful for tissue-
CC   specific gene therapy.
XX
SQ   Sequence 183 BP; 40 A; 53 C; 54 G; 36 T; 0 other;

  Query Match             79.1%;  Score 17.4;  DB 25;  Length 183;
  Best Local Similarity   94.7%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              |||||||||||||||||| |
Db        183 CGATCGGGGCGGGGCGAAC 165


RESULT 37
ABQ44940
ID   ABQ44940 standard; DNA; 722 BP.
XX
```

AC ABQ44940;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 31531.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 722 BP; 80 A; 93 C; 286 G; 263 T; 0 other;

Query Match 79.1%; Score 17.4; DB 24; Length 722;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCG 19
|||||.|||||||||||||
Db 514 ATTCGGTCGGGGCGGGGCG 532

RESULT 38
ABQ44941/c
ID ABQ44941 standard; DNA; 722 BP.
XX
AC ABQ44941;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 31532.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 722 BP; 263 A; 286 C; 93 G; 80 T; 0 other;

Query Match 79.1%; Score 17.4; DB 24; Length 722;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCG 19
|||||.|||||||||||||
Db 209 ATTCGGTCGGGGCGGGGCG 191

RESULT 39
ABQ43520
ID ABQ43520 standard; DNA; 747 BP.
XX
AC ABQ43520;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30111.
XX

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 747 BP; 80 A; 98 C; 302 G; 265 T; 2 other;

Query Match 79.1%; Score 17.4; DB 24; Length 747;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCG 19
||||| ||||||||||||
Db 539 ATTCGGTCGGGGCGGGGCG 557

RESULT 40
ABQ43521/c
ID ABQ43521 standard; DNA; 747 BP.
XX
AC ABQ43521;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30112.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 747 BP; 265 A; 302 C; 98 G; 80 T; 2 other;

Query Match 79.1%; Score 17.4; DB 24; Length 747;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCG 19
||||| ||||||||||||
Db 209 ATTCGGTCGGGGCGGGGCG 191

Search completed: February 18, 2004, 16:45:33
Job time : 173 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 14:54:35 ; Search time 32 Seconds
(without alignments)
303.451 Million cell updates/sec

Title: US-10-026-341A-2
Perfect score: 22
Sequence: 1 attcgatcggggcggggcgagc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	22	100.0	22	2	US-08-327-832-12	Sequence 12, Appl
	2	22	100.0	22	2	US-08-828-584-12	Sequence 12, Appl
	3	22	100.0	22	5	PCT-US94-05659-16	Sequence 16, Appl
	4	19	86.4	46	1	US-08-122-433-12	Sequence 12, Appl
	5	19	86.4	46	1	US-08-122-433-13	Sequence 13, Appl
c	6	18	81.8	46	1	US-08-122-433-12	Sequence 12, Appl
c	7	18	81.8	46	1	US-08-122-433-13	Sequence 13, Appl
c	8	16.8	76.4	2875	3	US-08-458-434A-4	Sequence 4, Appli
c	9	16.8	76.4	15144	3	US-08-458-434A-6	Sequence 6, Appli
c	10	16.4	74.5	379	1	US-09-591-383-5	Sequence 5, Appli
c	11	16.4	74.5	379	1	US-08-145-617-5	Sequence 5, Appli
	12	16.2	73.6	3997	4	US-09-345-236B-145	Sequence 145, App
c	13	16.2	73.6	3997	4	US-09-345-236B-146	Sequence 146, App
	14	16.2	73.6	9208	4	US-09-068-506-1	Sequence 1, Appli
c	15	15.8	71.8	31	1	US-08-153-563-4	Sequence 4, Appli
c	16	15.8	71.8	31	2	US-09-038-227-9	Sequence 9, Appli
c	17	15.8	71.8	31	2	US-08-460-507-4	Sequence 4, Appli
	18	15.8	71.8	944	2	US-08-786-606-4	Sequence 4, Appli
	19	15.8	71.8	1112	2	US-08-933-750C-97	Sequence 97, Appl
	20	15.8	71.8	1112	3	US-09-234-613-97	Sequence 97, Appl
	21	15.8	71.8	1114	4	US-09-690-454-39	Sequence 39, Appl
c	22	15.8	71.8	1919	1	US-07-991-587A-1	Sequence 1, Appli
c	23	15.8	71.8	1919	1	US-08-309-985-1	Sequence 1, Appli
c	24	15.8	71.8	3314	1	US-07-973-324A-5	Sequence 5, Appli
c	25	15.8	71.8	3314	1	US-08-343-380-5	Sequence 5, Appli
c	26	15.8	71.8	3314	3	US-09-072-435-5	Sequence 5, Appli
c	27	15.8	71.8	3314	3	US-09-072-917A-5	Sequence 5, Appli
c	28	15.6	70.9	327	4	US-09-252-991A-5577	Sequence 5577, Ap
c	29	15.6	70.9	730	2	US-08-743-637B-11	Sequence 11, Appl
c	30	15.6	70.9	730	3	US-08-526-840B-11	Sequence 11, Appl
	31	15.6	70.9	912	4	US-09-252-991A-5679	Sequence 5679, Ap
c	32	15.6	70.9	1007	3	US-08-836-500A-1	Sequence 1, Appli
c	33	15.6	70.9	1008	3	US-08-721-979A-13	Sequence 13, Appl
c	34	15.6	70.9	1008	4	US-09-654-289-13	Sequence 13, Appl
c	35	15.6	70.9	1008	4	US-09-582-876-13	Sequence 13, Appl
c	36	15.2	69.1	80	3	US-09-039-555B-4	Sequence 4, Appli
c	37	15.2	69.1	420	4	US-09-252-991A-10335	Sequence 10335, A
	38	15.2	69.1	894	4	US-09-252-991A-11338	Sequence 11338, A
	39	15.2	69.1	1335	4	US-09-252-991A-11357	Sequence 11357, A
	40	15.2	69.1	1506	4	US-09-252-991A-10637	Sequence 10637, A
	41	15.2	69.1	1605	4	US-09-252-991A-10729	Sequence 10729, A
c	42	15.2	69.1	4284	4	US-09-252-991A-10434	Sequence 10434, A
	43	15.2	69.1	9704	4	US-09-814-951A-3	Sequence 3, Appli
	44	15.2	69.1	44377	2	US-08-804-227C-7	Sequence 7, Appli
	45	15.2	69.1	44377	2	US-08-804-198-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-327-832-12
; Sequence 12, Application US/08327832
; Patent No. 5840832
;  GENERAL INFORMATION:
;    APPLICANT:  Ono, Santa J.
;    APPLICANT:  Strominger, Jack L.
;    TITLE OF INVENTION:  Transcription Factor Regulating MHC
;    TITLE OF INVENTION:  Expression, cDNA and Genomic Clones Encoding Same and
;    TITLE OF INVENTION:  Retroviral Expression Contructs Thereof
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington, D.C.
;      STATE:  District of Columbia
;      COUNTRY:  U.S.A.
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/327,832
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Posorske, Laurence H.
;      REGISTRATION NUMBER:  34,698
;      REFERENCE/DOCKET NUMBER:  1107.46362
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  20-2 508-9153
;      TELEFAX:  202 508-9299
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  homo sapiens
US-08-327-832-12

  Query Match           100.0%;  Score 22;  DB 2;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 0.46;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
```

```
                    ||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 2
US-08-828-584-12
; Sequence 12, Application US/08828584
; Patent No. 5908762
;  GENERAL INFORMATION:
;    APPLICANT:  Ono, Santa J.
;    APPLICANT:  Strominger, Jack L.
;    TITLE OF INVENTION:  Transcription Factor Regulating MHC
;    TITLE OF INVENTION:  Expression, cDNA and Genomic Clones Encoding Same and
;    TITLE OF INVENTION:  Retroviral Expression Contructs Thereof
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington, D.C.
;      STATE:  District of Columbia
;      COUNTRY:  U.S.A.
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/828,584
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Posorske, Laurence H.
;      REGISTRATION NUMBER:  34,698
;      REFERENCE/DOCKET NUMBER:  1107.46362
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  20-2 508-9153
;      TELEFAX:  202 508-9299
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  homo sapiens
US-08-828-584-12

  Query Match             100.0%;  Score 22;  DB 2;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
                ||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 3
PCT-US94-05659-16
; Sequence 16, Application PC/TUS9405659
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  TNF' RESPONSIVE ELEMENT, TNF'-INDUCED DNA-BINDING
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/05659
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  FDC93-01 FF
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US94-05659-16

  Query Match             100.0%;  Score 22;  DB 5;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
                ||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 4
US-08-122-433-12
; Sequence 12, Application US/08122433
; Patent No. 5683985
;  GENERAL INFORMATION:
;    APPLICANT:  Chu, Barbara C.F.
;    APPLICANT:  Orgel, Leslie
;    TITLE OF INVENTION:  OLIGODEOXYNUCLEOTIDES AND
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES USEFUL AS DECOYS FOR PROTEINS WHICH
;    TITLE OF INVENTION:  SELECTIVELY BIND TO DEFINED DNA SEQUENCES
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/122,433
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/687,337
;      FILING DATE:  18-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P31 9308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-1995
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  46 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-122-433-12

  Query Match             86.4%;  Score 19;  DB 1;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 8.3;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||
Db         28 TCGATCGGGGCGGGGCGAG 46


RESULT 5
US-08-122-433-13
; Sequence 13, Application US/08122433
; Patent No. 5683985
;  GENERAL INFORMATION:
;    APPLICANT:  Chu, Barbara C.F.
;    APPLICANT:  Orgel, Leslie
;    TITLE OF INVENTION:  OLIGODEOXYNUCLEOTIDES AND
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES USEFUL AS DECOYS FOR PROTEINS WHICH
;    TITLE OF INVENTION:  SELECTIVELY BIND TO DEFINED DNA SEQUENCES
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/122,433
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/687,337
;      FILING DATE:  18-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P31 9308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-1995
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  46 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  other nucleic acid
US-08-122-433-13

  Query Match             86.4%;  Score 19;  DB 1;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 8.3;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||
Db         28 TCGATCGGGGCGGGGCGAG 46


RESULT 6
US-08-122-433-12/c
; Sequence 12, Application US/08122433
; Patent No. 5683985
;  GENERAL INFORMATION:
;    APPLICANT:  Chu, Barbara C.F.
;    APPLICANT:  Orgel, Leslie
;    TITLE OF INVENTION:  OLIGODEOXYNUCLEOTIDES AND
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES USEFUL AS DECOYS FOR PROTEINS WHICH
;    TITLE OF INVENTION:  SELECTIVELY BIND TO DEFINED DNA SEQUENCES
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/122,433
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/687,337
;      FILING DATE:  18-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P31 9308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-1995
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  46 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-122-433-12

  Query Match             81.8%;  Score 18;  DB 1;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||
Db         24 GATCGGGGCGGGGCGAGC 7


RESULT 7
US-08-122-433-13/c
; Sequence 13, Application US/08122433
; Patent No. 5683985
;  GENERAL INFORMATION:
;    APPLICANT:  Chu, Barbara C.F.
;    APPLICANT:  Orgel, Leslie
;    TITLE OF INVENTION:  OLIGODEOXYNUCLEOTIDES AND
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES USEFUL AS DECOYS FOR PROTEINS WHICH
;    TITLE OF INVENTION:  SELECTIVELY BIND TO DEFINED DNA SEQUENCES
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/122,433
;       FILING DATE:  22-SEP-1993
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/687,337
;       FILING DATE:  18-APR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Reiter, Stephen E.
;       REGISTRATION NUMBER:  31,192
;       REFERENCE/DOCKET NUMBER:  P31 9308
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-546-1995
;       TELEFAX:  619-546-9392
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  46 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  other nucleic acid
US-08-122-433-13

  Query Match             81.8%;  Score 18;  DB 1;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||
Db         24 GATCGGGGCGGGGCGAGC 7


RESULT 8
US-08-458-434A-4/c
; Sequence 4, Application US/08458434A
; Patent No. 6083690
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Stephen E.
;    APPLICANT:  Mundy M.D., Gregory R.
;    APPLICANT:  Gosh-Choudhury Ph.D., Nandini
;    APPLICANT:  Feng Ph.D., Jian Q.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR IDENTIFYING
;    TITLE OF INVENTION:  OSTEOGENIC AGENTS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  James C. Weseman, Esq.
;      STREET:  401 B. Street, Suite 1700
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,434A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weseman, James C.
;      REGISTRATION NUMBER:  30,507
;      REFERENCE/DOCKET NUMBER:  P00060US0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 699-3604
;      TELEFAX:  619-236-1048
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2875 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-458-434A-4

  Query Match             76.4%;  Score 16.8;  DB 3;  Length 2875;
  Best Local Similarity   90.0%;  Pred. No. 60;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              | || ||||||||||||||||
Db       1805 TGGAGCGGGGCGGGGCGAGC 1786


RESULT 9
US-08-458-434A-6/c
; Sequence 6, Application US/08458434A
; Patent No. 6083690
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Stephen E.
;    APPLICANT:  Mundy M.D., Gregory R.
;    APPLICANT:  Gosh-Choudhury Ph.D., Nandini
;    APPLICANT:  Feng Ph.D., Jian Q.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR IDENTIFYING
;    TITLE OF INVENTION:  OSTEOGENIC AGENTS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  James C. Weseman, Esq.
;      STREET:  401 B. Street, Suite 1700
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,434A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weseman, James C.
;      REGISTRATION NUMBER:  30,507
;      REFERENCE/DOCKET NUMBER:  P00060US0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 699-3604
;      TELEFAX:  619-236-1048
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15144 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-458-434A-6

  Query Match             76.4%;  Score 16.8;  DB 3;  Length 15144;
  Best Local Similarity   90.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              | || ||||||||||||||||
Db       1805 TGGAGCGGGGCGGGGCGAGC 1786


RESULT 10
US-09-591-383-5/c
; Sequence 5, Application US/09591383
; Patent No. RE37984
;   GENERAL INFORMATION:
;        APPLICANT: Jackle, Herbert
```

```
;                 Tautz, Diethard
;       TITLE OF INVENTION: PROCESS FOR ANALYZING LENGTH
;                           POLYMORPHISMS IN DNA REGIONS
;       NUMBER OF SEQUENCES: 6
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
;            STREET: 301 N. Washington Street, P.O. Box 747
;            CITY: Falls Church
;            STATE: Virginia
;            COUNTRY: United States of America
;            ZIP: 22046
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/591,383
;            FILING DATE: 09-Jun-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 07/681,494
;            FILING DATE: 10-JUN-1991
;            APPLICATION NUMBER: DE P3834636.2
;            FILING DATE: 11-OCT-1988
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svensson, Leonard R.
;            REGISTRATION NUMBER: 30,330
;            REFERENCE/DOCKET NUMBER: 147-122PCT
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 703-241-1300
;            TELEFAX: 703-241-2848
;            TELEX: 248345
;   INFORMATION FOR SEQ ID NO: 5:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 379 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-591-383-5

  Query Match             74.5%;  Score 16.4;  DB 1;  Length 379;
  Best Local Similarity   94.4%;  Pred. No. 96;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          5 GATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db         64 GATTGGGGCGGGGCGAGC 47


RESULT 11
US-08-145-617-5/c
; Sequence 5, Application US/08145617
; Patent No. 5766847
;  GENERAL INFORMATION:
;    APPLICANT:  Jackle, Herbert
;    APPLICANT:  Tautz, Diethard
;    TITLE OF INVENTION:  PROCESS FOR ANALYZING LENGTH
;    TITLE OF INVENTION:  POLYMORPHISMS IN DNA REGIONS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  301 N. Washington Street, P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  United States of America
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/145,617
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/681,494
;      FILING DATE:  10-JUN-1991
;      APPLICATION NUMBER:  DE P3834636.2
;      FILING DATE:  11-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svensson, Leonard R.
;      REGISTRATION NUMBER:  30,330
;      REFERENCE/DOCKET NUMBER:  147-122PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-241-1300
;      TELEFAX:  703-241-2848
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  379 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-145-617-5

  Query Match             74.5%;  Score 16.4;  DB 1;  Length 379;
  Best Local Similarity   94.4%;  Pred. No. 96;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          5 GATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db         64 GATTGGGGCGGGGCGAGC 47


RESULT 12
US-09-345-236B-145
; Sequence 145, Application US/09345236B
; Patent No. 6521454
; GENERAL INFORMATION:
;  APPLICANT: Becnel, James J.
;  APPLICANT:  Tukuo, Fukuda
;  APPLICANT:  Moser, Bettina
;  APPLICANT:  Cockburn, Andrew
;  APPLICANT:  White, Susan E.
;  APPLICANT:  Undeen, Albert H.
;  TITLE OF INVENTION: No. 6521454el Baculoviruses,  Insecticidal
;  TITLE OF INVENTION:  Compositions, and Methods for Control of Invertebrates
;  FILE REFERENCE: 21042.0004
;  CURRENT APPLICATION NUMBER: US/09/345,236B
;  CURRENT FILING DATE:  1999-06-30
;  NUMBER OF SEQ ID NOS: 148
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 145
;   LENGTH: 3997
;   TYPE: DNA
;   ORGANISM: mosquito baculovirus
US-09-345-236B-145

  Query Match             73.6%;  Score 16.2;  DB 4;  Length 3997;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              ||||| ||||| |||| ||||
Db       2827 ATTCGTTCGGGTCGGGTCGAG 2847


RESULT 13
US-09-345-236B-146/c
; Sequence 146, Application US/09345236B
; Patent No. 6521454
```

```
; GENERAL INFORMATION:
;  APPLICANT: Becnel, James J.
;  APPLICANT:  Tukuo, Fukuda
;  APPLICANT:  Moser, Bettina
;  APPLICANT:  Cockburn, Andrew
;  APPLICANT:  White, Susan E.
;  APPLICANT:  Undeen, Albert H.
;  TITLE OF INVENTION: No. 6521454e1 Baculoviruses,  Insecticidal
;  TITLE OF INVENTION:  Compositions, and Methods for Control of Invertebrates
;  FILE REFERENCE: 21042.0004
;  CURRENT APPLICATION NUMBER: US/09/345,236B
;  CURRENT FILING DATE:  1999-06-30
;  NUMBER OF SEQ ID NOS: 148
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 146
;   LENGTH: 3997
;   TYPE: DNA
;   ORGANISM: mosquito baculovirus
US-09-345-236B-146

  Query Match             73.6%;  Score 16.2;  DB 4;  Length 3997;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              ||||| ||||| |||| ||||
Db       1171 ATTCGTTCGGGTCGGGTCGAG 1151


RESULT 14
US-09-068-506-1
; Sequence 1, Application US/09068506A
; Patent No. 6569618
; GENERAL INFORMATION:
;  APPLICANT: YASUE, Hirofumi
;  APPLICANT: YOSHIMURA, Kumamoto
;  TITLE OF INVENTION: DIAGNOSIS OF DISEASES ASSOCIATED WITH CORONARY
;  TITLE OF INVENTION: TWITCHING
;  FILE REFERENCE: 0032-245P
;  CURRENT APPLICATION NUMBER: US/09/068,506A
;  CURRENT FILING DATE: 1998-07-10
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 9208
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: nnnnnnnnnn = Intervening sequences of introns
US-09-068-506-1

  Query Match             73.6%;  Score 16.2;  DB 4;  Length 9208;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              ||  ||| ||||||||||||||
Db       1489 ATGGGATAGGGGCGGGGCGAG 1509


RESULT 15
US-08-153-563-4/c
; Sequence 4, Application US/08153563
; Patent No. 5693506
;  GENERAL INFORMATION:
;    APPLICANT:  Rodriquez, Raymond L.
;    TITLE OF INVENTION:  PROCESS FOR PROTEIN PRODUCTION
;    TITLE OF INVENTION:  IN PLANTS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94105-1493
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/153,563
;      FILING DATE:  16-NOV-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weber, Kenneth A.
;      REGISTRATION NUMBER:  31,677
;      REFERENCE/DOCKET NUMBER:  2307E-515
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..31
;      OTHER INFORMATION:  /standard_name= "31 bp RAmy3E"
US-08-153-563-4

  Query Match             71.8%;  Score 15.8;  DB 1;  Length 31;
  Best Local Similarity   89.5%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              |||||| ||||| ||||||
Db         23 CGATCGAGGCGGCGCGAGC 5


RESULT 16
US-09-038-227-9/c
; Sequence 9, Application US/09038227
; Patent No. 5917029
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Su-May
;    TITLE OF INVENTION:  SUGAR-RESPONSIVE ENHANCERS
;    TITLE OF INVENTION:  IN ALPHA-AMYLASE GENES
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/038,227
;      FILING DATE:  11-MAR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsao, Y. Rocky
;      REGISTRATION NUMBER:  34,053
;      REFERENCE/DOCKET NUMBER:  05228/031001
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  31 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
US-09-038-227-9

  Query Match             71.8%;  Score 15.8;  DB 2;  Length 31;
  Best Local Similarity   89.5%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              |||||| ||||| ||||||
Db         23 CGATCGAGGCGGCGCGAGC 5


RESULT 17
US-08-460-507-4/c
; Sequence 4, Application US/08460507
; Patent No. 5994628
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  PROCESS FOR PROTEIN PRODUCTION
;    TITLE OF INVENTION:  IN PLANTS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,507
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,563
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fabian, Gary R.
;      REGISTRATION NUMBER:  33,875
;      REFERENCE/DOCKET NUMBER:  2000-0452.41
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..31
;      OTHER INFORMATION:  /standard_name= "31 bp RAmy3E"
US-08-460-507-4

  Query Match             71.8%;  Score 15.8;  DB 2;  Length 31;
  Best Local Similarity   89.5%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              |||||| ||||| ||||||
Db         23 CGATCGAGGCGGCGCGAGC 5


RESULT 18
US-08-786-606-4
; Sequence 4, Application US/08786606
; Patent No. 5861495
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Coleman, Roger
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN ZINC-BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/786,606
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy RJ
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0173 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  944 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-786-606-4

  Query Match             71.8%;  Score 15.8;  DB 2;  Length 944;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| |||||||||||| ||
Db        145 CGAGCGGGGCGGGGCGGGC 163


RESULT 19
US-08-933-750C-97
; Sequence 97, Application US/08933750C
; Patent No. 5932442
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
```

```
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,750C
;      FILING DATE:  September 23, 1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0356 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  97:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1112 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  TESTNOT07
;      CLONE:  3217567
US-08-933-750C-97

  Query Match             71.8%;  Score 15.8;  DB 2;  Length 1112;
  Best Local Similarity   89.5%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| |||||||||||| ||
Db        147 CGAGCGGGGCGGGGCGGGC 165


RESULT 20
US-09-234-613-97
; Sequence 97, Application US/09234613
; Patent No. 6132973
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/234,613
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,750
;      FILING DATE:  September 23, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0356 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  97:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1112 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  TESTNOT07
;      CLONE:  3217567
US-09-234-613-97

  Query Match             71.8%;  Score 15.8;  DB 3;  Length 1112;
  Best Local Similarity   89.5%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| |||||||||||| ||
Db        147 CGAGCGGGGCGGGGCGGGC 165


RESULT 21
US-09-690-454-39
; Sequence 39, Application US/09690454
; Patent No. 6531447
; GENERAL INFORMATION:
;  APPLICANT: Steven M. Ruben, et al.
;  TITLE OF INVENTION: 32 Human Secreted Proteins
;  FILE REFERENCE: PZ006P1
;  CURRENT APPLICATION NUMBER: US/09/690,454
;  CURRENT FILING DATE:  2000-10-18
;  PRIOR APPLICATION NUMBER: 09/189,144
;  PRIOR FILING DATE: 1998-11-10
;  PRIOR APPLICATION NUMBER: 60/044,039
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,093
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,190
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/050,935
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,101
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/048,356
;  PRIOR FILING DATE: May 30, 1997
;  PRIOR APPLICATION NUMBER: 60/056,250
;  PRIOR FILING DATE: August 29, 1997
;  PRIOR APPLICATION NUMBER: 60/056,296
;  PRIOR FILING DATE: August 29, 1997
;  PRIOR APPLICATION NUMBER: 60/056,293
;  PRIOR FILING DATE: August 29, 1997
;  NUMBER OF SEQ ID NOS: 229
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
;   LENGTH: 1114
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-690-454-39

  Query Match             71.8%;  Score 15.8;  DB 4;  Length 1114;
  Best Local Similarity   89.5%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| |||||||||||| ||
Db         96 CGAGCGGGGCGGGGCGGGC 114


RESULT 22
US-07-991-587A-1/c
; Sequence 1, Application US/07991587A
; Patent No. 5384249
;  GENERAL INFORMATION:
;    APPLICANT:  Sasaki, Katsutoshi
;    APPLICANT:  Watanabe, Etsuyo
;    APPLICANT:  Nishi, Tatsunari
;    APPLICANT:  Sekine, Susumu
;    APPLICANT:  Hanai, No. 5384249uo
;    APPLICANT:  Hasegawa, Mamoru
;    TITLE OF INVENTION:  '2  3 Sialyltransferase
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fitzpatrick, Cella, Harper & Scinto
;      STREET:  277 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10172
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44mb
;      COMPUTER:  IBM PC
;      OPERATING SYSTEM:  Dos 3.3
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/991,587A
;      FILING DATE:  19930526
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP-333661/1991
;      APPLICATION NUMBER:  JP-091044/1992
;      FILING DATE:  17-12-1991
;      FILING DATE:  10-04-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lawrence S. Perry
;      REGISTRATION NUMBER:  31,865
;      REFERENCE/DOCKET NUMBER:  1580.2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-758-2400
;      TELEFAX:  212-758-2982
;      TELEX:  236262
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1919
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  human
;      CELL LINE:  TYH cell
;      CELL TYPE:  histiocytoma cell
US-07-991-587A-1

  Query Match             71.8%;  Score 15.8;  DB 1;  Length 1919;
  Best Local Similarity   89.5%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
                | || ||||||||||||||
Db         59 TGGAGCGGGGCGGGGCGAG 41


RESULT 23
US-08-309-985-1/c
; Sequence 1, Application US/08309985
; Patent No. 5494790
;  GENERAL INFORMATION:
;    APPLICANT:  Sasaki, Katsutoshi
;    APPLICANT:  Watanabe, Etsuyo
;    APPLICANT:  Nishi, Tatsunari
;    APPLICANT:  Sekine, Susumu
;    APPLICANT:  Hanai, No. 5494790uo
;    APPLICANT:  Hasegawa, Mamoru
;    TITLE OF INVENTION:  '2  3 Sialyltransferase
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fitzpatrick, Cella, Harper & Scinto
;      STREET:  277 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10172
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44mb
;      COMPUTER:  IBM PC
;      OPERATING SYSTEM:  Dos 3.3
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/309,985
;      FILING DATE:  20-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/991,587
;      FILING DATE:  16-12-1992
;      APPLICATION NUMBER:  JP-333661/1991
;      APPLICATION NUMBER:  JP-091044/1992
;      FILING DATE:  17-12-1991
;      FILING DATE:  10-04-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lawrence S. Perry
;      REGISTRATION NUMBER:  31,865
;      REFERENCE/DOCKET NUMBER:  1580.2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-758-2400
;      TELEFAX:  212-758-2982
;      TELEX:  236262
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1919
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  human
;      CELL LINE:  TYH cell
;      CELL TYPE:  histiocytoma cell
US-08-309-985-1

  Query Match             71.8%;  Score 15.8;  DB 1;  Length 1919;
  Best Local Similarity   89.5%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
                | || ||||||||||||||
Db         59 TGGAGCGGGGCGGGGCGAG 41


RESULT 24
US-07-973-324A-5/c
; Sequence 5, Application US/07973324A
```

```
; Patent No. 5460952
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Su-May
;    APPLICANT:  Liu, Li-Fei
;    TITLE OF INVENTION:  Gene Expression System Comprising the
;    TITLE OF INVENTION:  Promoter Region of the Alpha-Amylase Genes
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/973,324A
;      FILING DATE:  04-NOV-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Borun, Michael F.
;      REGISTRATION NUMBER:  25447
;      REFERENCE/DOCKET NUMBER:  31149
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Rice (Oryzae sativa)
;      STRAIN:  CV. M202
;    IMMEDIATE SOURCE:
;      LIBRARY:  (EMBL) genomic
;      CLONE:  '-Amy8-C
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1152..1241, 1385..2323, 2409..2690)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  join(1227..1241, 1385..2323, 2409..2690)
;    PUBLICATION INFORMATION:
;      AUTHORS:  Yu et al., Su-May
;      TITLE:  Regulation of '-amylase-encoding gene expression
;      TITLE:  in germinating seeds and cultured cells of rice
;      JOURNAL:  Gene
;      VOLUME:  in press
US-07-973-324A-5

  Query Match             71.8%;  Score 15.8;  DB 1;  Length 3314;
  Best Local Similarity   89.5%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              |||||| ||||| ||||||
Db        782 CGATCGAGGCGGCGCGAGC 764


RESULT 25
US-08-343-380-5/c
; Sequence 5, Application US/08343380
; Patent No. 5712112
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Su-May
;    APPLICANT:  Liu, Li-Fei
;    TITLE OF INVENTION:  Gene Expression System Comprising the
;    TITLE OF INVENTION:  Promoter Region of the Alpha-Amylase Genes
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/343,380
;      FILING DATE:  22-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/973,324
;      FILING DATE:  04-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Borun, Michael F.
;      REGISTRATION NUMBER:  25447
;      REFERENCE/DOCKET NUMBER:  31149
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Rice (Oryzae sativa)
;      STRAIN:  CV. M202
;    IMMEDIATE SOURCE:
;      LIBRARY:  (EMBL) genomic
;      CLONE:  '-Amy8-C
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1152..1241, 1385..2323, 2409..2690)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  join(1227..1241, 1385..2323, 2409..2690)
;    PUBLICATION INFORMATION:
;      AUTHORS:  Yu et al., Su-May
;      TITLE:  Regulation of '-amylase-encoding gene expression
;      TITLE:  in germinating seeds and cultured cells of rice
;      JOURNAL:  Gene
;      VOLUME:  in press
US-08-343-380-5

  Query Match             71.8%;  Score 15.8;  DB 1;  Length 3314;
  Best Local Similarity   89.5%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              |||||| ||||| ||||||
Db        782 CGATCGAGGCGGCGCGAGC 764


RESULT 26
US-09-072-435-5/c
; Sequence 5, Application US/09072435
; Patent No. 6215051
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Su-May
```

```
;    APPLICANT:  Liu, Li-Fei
;    APPLICANT:  Chan, Ming-Tsair
;    TITLE OF INVENTION:  GENE EXPRESSION SYSTEM COMPRISING THE
;    TITLE OF INVENTION:  PROMOTER REGION OF THE ALPHA-AMYLASE GENES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive/6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/072,435
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/639,792
;      FILING DATE:  29-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/973,324
;      FILING DATE:  04-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,153
;      REFERENCE/DOCKET NUMBER:  28123/34274
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Rice (Oryzae sativa)
;      STRAIN:  CV. M202
;    IMMEDIATE SOURCE:
;      LIBRARY:  (EMBL) genomic
;      CLONE:  '-Amy8-C
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1152..1241, 1385..2323, 2409..2690)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  join(1227..1241, 1385..2323, 2409..2690)
;    PUBLICATION INFORMATION:
;      AUTHORS:  Yu et al., Su-May
;      TITLE:  Regulation of '-amylase-encoding gene expression
;      TITLE:  in germinating seeds and cultured cells of rice
;      JOURNAL:  Gene
;      VOLUME:  in press
US-09-072-435-5

  Query Match             71.8%;  Score 15.8;  DB 3;  Length 3314;
  Best Local Similarity   89.5%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              |||||| ||||| ||||||
Db        782 CGATCGAGGCGGCGCGAGC 764


RESULT 27
US-09-072-917A-5/c
; Sequence 5, Application US/09072917A
; Patent No. 6288302
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Su-May
;    APPLICANT:  Liu, Li-Fei
;    APPLICANT:  Chan, Ming-Tsair
;    TITLE OF INVENTION:  Application of Alpha-Amylase Gene
;    TITLE OF INVENTION:  Promoter and Signal Sequence in the Production of
; Patent No. 6288302
;    TITLE OF INVENTION:  Recombinant Proteins in Transgenic Plants and Transgenic
;    TITLE OF INVENTION:  Plant Seeds
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive/6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/072,917A
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/509,962
;      FILING DATE:  01-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,153
;      REFERENCE/DOCKET NUMBER:  28123/34257
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Rice (Oryzae sativa)
;      STRAIN:  CV. M202
;    IMMEDIATE SOURCE:
;      LIBRARY:  (EMBL) genomic
;      CLONE:  alpha-Amy8-C
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1152..1241, 1385..2323, 2409..2690)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  join(1227..1241, 1385..2323, 2409..2690)
;    PUBLICATION INFORMATION:
;      AUTHORS:  Yu et al., Su-May
;      TITLE:  Regulation of alpha-amylase-encoding gene expression
;      TITLE:  in germinating seeds and cultured cells of rice
;      JOURNAL:  Gene
;      VOLUME:  in press
US-09-072-917A-5

  Query Match             71.8%;  Score 15.8;  DB 3;  Length 3314;
  Best Local Similarity   89.5%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              |||||| ||||| ||||||
Db        782 CGATCGAGGCGGCGCGAGC 764
```

```
RESULT 28
US-09-252-991A-5577/c
; Sequence 5577, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 5577
;   LENGTH: 327
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5577

  Query Match             70.9%;  Score 15.6;  DB 4;  Length 327;
  Best Local Similarity   81.8%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              |||||| || ||| |||||| |
Db        282 ATTCGAACGTGGCCGGGCGACC 261


RESULT 29
US-08-743-637B-11/c
; Sequence 11, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  PICARD, Francois J.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;    TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;    TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
;    TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 EAST WISCONSIN AVENUE
;      CITY:  MILWAUKEE
;      STATE:  WISCONSIN
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/743,637B
;      FILING DATE:  04-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,840
;      FILING DATE:  11-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414)277-5591
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  730 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Klebsiella pneumoniae
US-08-743-637B-11

  Query Match             70.9%;  Score 15.6;  DB 2;  Length 730;
  Best Local Similarity   81.8%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              | |||||| |||| | ||||||
Db        686 AATCGATCAGGGCAGCGCGAGC 665


RESULT 30
US-08-526-840B-11/c
; Sequence 11, Application US/08526840B
; Patent No. 6001564
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIFIC AND UNIVERSAL PROBES AND
;    TITLE OF INVENTION:  AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
;    TITLE OF INVENTION:  COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
;    TITLE OF INVENTION:  FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
;    NUMBER OF SEQUENCES:  177
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/526,840B
;      FILING DATE:  11-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/304,732
;      FILING DATE:  12-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414) 277-5591
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  730 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Klebsiella pneumoniae
US-08-526-840B-11

  Query Match             70.9%;  Score 15.6;  DB 3;  Length 730;
  Best Local Similarity   81.8%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
```

```
                  | |||||| |||| | ||||||
Db        686 AATCGATCAGGGCAGCGCGAGC 665


RESULT 31
US-09-252-991A-5679
; Sequence 5679, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 5679
;   LENGTH: 912
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5679

  Query Match             70.9%;  Score 15.6;  DB 4;  Length 912;
  Best Local Similarity   81.8%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              |||||| || ||| |||||| |
Db        165 ATTCGAACGTGGCCGGGCGACC 186


RESULT 32
US-08-836-500A-1/c
; Sequence 1, Application US/08836500A
; Patent No. 6197929
;  GENERAL INFORMATION:
;    APPLICANT:  Binz, Hans
;    APPLICANT:  Baussant, Thierry
;    APPLICANT:  Haeuw, Jean-Francois
;    APPLICANT:  Nguyen Ngoc, Thien
;    TITLE OF INVENTION:  Carrier Protein Having an Adjuvant
;    TITLE OF INVENTION:  Effect, Immunogenic Complex Containing It, Process for
;    TITLE OF INVENTION:  Their Preparation, Nucleotide Sequence and Vaccines
; Patent No. 6197929
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rockey, Milnamow & Katz, Ltd.
;      STREET:  180 N. Stetson, 2 Prudential Plaza, Suite
;      STREET:  4700
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  U.S.A.
;      ZIP:  60601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,500A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Katz, Martin L.
;      REGISTRATION NUMBER:  25,011
;      REFERENCE/DOCKET NUMBER:  PIE1514P0180US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-616-5400
;      TELEFAX:  312-616-5460
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1007 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1007
US-08-836-500A-1

  Query Match             70.9%;  Score 15.6;  DB 3;  Length 1007;
  Best Local Similarity   81.8%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              | |||||| |||| | ||||||
Db        934 AATCGATCAGGGCAGCGCGAGC 913


RESULT 33
US-08-721-979A-13/c
; Sequence 13, Application US/08721979A
; Patent No. 6113911
;  GENERAL INFORMATION:
;    APPLICANT:  Binz, Hans
;    APPLICANT:  N'Guyen, Ngoc Thien
;    APPLICANT:  Baussant, Thierry
;    APPLICANT:  Trudel, Michel
;    TITLE OF INVENTION:  PEPTIDE FRAGMENT OF RESPIRATORY
;    TITLE OF INVENTION:  SYNCYTIAL VIRUS PROTEIN G, IMMUNOGENIC AGENT, PHARMACEUTICAL
;    TITLE OF INVENTION:  COMPOSITION CONTAINING IT AND PREPARATION PROCESS
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gordon W. Hueschen
;      STREET:  715 The "H" Bldg., 310 East Michigan
;      STREET:  Avenue
;      CITY:  Kalamazoo
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  49007
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      SOFTWARE:
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/721,979A
;      FILING DATE:  October 4, 1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 94 04009
;      FILING DATE:  06-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hueschen, Gordon W.
;      REGISTRATION NUMBER:  16,157
;      REFERENCE/DOCKET NUMBER:  PF57PCTUS/dln
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  616-382-0030
;      TELEFAX:  616-382-2030
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1008 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1008
;      OTHER INFORMATION:  /note= "name : P40"
```

```
US-08-721-979A-13

  Query Match             70.9%;  Score 15.6;  DB 3;  Length 1008;
  Best Local Similarity   81.8%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              | |||||| |||| | ||||||
Db        934 AATCGATCAGGGCAGCGCGAGC 913


RESULT 34
US-09-654-289-13/c
; Sequence 13, Application US/09654289
; Patent No. 6410030
;  GENERAL INFORMATION:
;    APPLICANT:  Binz, Hans
;    APPLICANT:  N'Guyen, Ngoc Thien
;    APPLICANT:  Baussant, Thierry
;    APPLICANT:  Trudel, Michel
;    TITLE OF INVENTION:  PEPTIDE FRAGMENT OF RESPIRATORY
;    TITLE OF INVENTION:  SYNCYTIAL VIRUS PROTEIN G, IMMUNOGENIC AGENT,PHARMACEUTICAL
;    TITLE OF INVENTION:  COMPOSITION CONTAINING IT AND PREPARATION PROCESS
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gordon W. Hueschen
;      STREET:  715 The "H" Bldg., 310 East Michigan
;      STREET:  Avenue
;      CITY:  Kalamazoo
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  49007
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      SOFTWARE:
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/654,289
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/721,979
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hueschen, Gordon W.
;      REGISTRATION NUMBER:  16,157
;      REFERENCE/DOCKET NUMBER:  PF57PCTUS/dln
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  616-382-0030
;      TELEFAX:  616-382-2030
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1008 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1008
;      OTHER INFORMATION:  /note= "name : P40"
US-09-654-289-13

  Query Match             70.9%;  Score 15.6;  DB 4;  Length 1008;
  Best Local Similarity   81.8%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              | |||||| |||| | ||||||
Db        934 AATCGATCAGGGCAGCGCGAGC 913


RESULT 35
US-09-582-876-13/c
; Sequence 13, Application US/09582876
; Patent No. 6537556
;  GENERAL INFORMATION:
;      APPLICANT: Binz, Hans
;                 N'Guyen, Ngoc Thien
;                 Baussant, Thierry
;                 Trudel, Michel
;      TITLE OF INVENTION: PEPTIDE FRAGMENT OF RESPIRATORY
;                          SYNCYTIAL VIRUS PROTEIN G, IMMUNOGENIC AGENT,PHARMACEUTIC
;                          COMPOSITION CONTAINING IT AND PREPARATION PROCESS
;      NUMBER OF SEQUENCES: 75
;      CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Gordon W. Hueschen
;           STREET: 715 The "H" Bldg., 310 East Michigan
;                   Avenue
;           CITY: Kalamazoo
;           STATE: MI
;           COUNTRY: USA
;           ZIP: 49007
;      COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
;                     (EPO)
;      CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/582,876
;           FILING DATE: 30-Jun-2000
;           CLASSIFICATION: <Unknown>
;      PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: 08/721,979
;           FILING DATE: <Unknown>
;      ATTORNEY/AGENT INFORMATION:
;           NAME: Hueschen, Gordon W.
;           REGISTRATION NUMBER: 16,157
;           REFERENCE/DOCKET NUMBER: PF57PCTUS/dln
;      TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 616-382-0030
;           TELEFAX: 616-382-2030
;  INFORMATION FOR SEQ ID NO: 13:
;      SEQUENCE CHARACTERISTICS:
;           LENGTH: 1008 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;      MOLECULE TYPE: peptide
;      FEATURE:
;           NAME/KEY:  CDS
;           LOCATION: 1..1008
;           OTHER INFORMATION: /note= "name : P40"
;      SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-582-876-13

  Query Match             70.9%;  Score 15.6;  DB 4;  Length 1008;
  Best Local Similarity   81.8%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              | |||||| |||| | ||||||
Db        934 AATCGATCAGGGCAGCGCGAGC 913


RESULT 36
US-09-039-555B-4/c
; Sequence 4, Application US/09039555B
; Patent No. 6033856
;  GENERAL INFORMATION:
;    APPLICANT:  Koerner, Kathrin
;    APPLICANT:  Mueller, Rolf
;    APPLICANT:  Sadlacek, Hans-Harald
```

```
;    TITLE OF INVENTION:  PROMOTER OF THE CDC25B GENE, ITS
;    TITLE OF INVENTION:  PREPARATION AND USE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/039,555B
;      FILING DATE:  16-MAR-1998
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19710643.9
;      FILING DATE:  14-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  016779/0131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "oligonucleotide"
US-09-039-555B-4

  Query Match             69.1%;  Score 15.2;  DB 3;  Length 80;
  Best Local Similarity   85.0%;  Pred. No. 3.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              || |||| ||||||||| ||
Db         34 TCCATCGTGGCGGGGCGGGC 15


RESULT 37
US-09-252-991A-10335/c
; Sequence 10335, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10335
;   LENGTH: 420
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10335

  Query Match             69.1%;  Score 15.2;  DB 4;  Length 420;
  Best Local Similarity   85.0%;  Pred. No. 3.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||| ||| ||||||||| |
Db        243 TCGGTCGCGGCGGGGCGATC 224


RESULT 38
US-09-252-991A-11338
; Sequence 11338, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11338
;   LENGTH: 894
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11338

  Query Match             69.1%;  Score 15.2;  DB 4;  Length 894;
  Best Local Similarity   85.0%;  Pred. No. 3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||   |||||||||
Db        675 TCGATCGGCCGGGGGCGAGC 694


RESULT 39
US-09-252-991A-11357
; Sequence 11357, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11357
;   LENGTH: 1335
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11357

  Query Match             69.1%;  Score 15.2;  DB 4;  Length 1335;
  Best Local Similarity   85.0%;  Pred. No. 2.9e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||   |||||||||
Db       1046 TCGATCGGCCGGGGGCGAGC 1065


RESULT 40
US-09-252-991A-10637
; Sequence 10637, Application US/09252991A
; Patent No. 6551795
```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10637
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10637

Query Match 69.1%; Score 15.2; DB 4; Length 1506;
Best Local Similarity 85.0%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||| ||| |||||||||| |
Db        435 TCGGTCGCGGCGGGGCGATC 454
```

Search completed: February 18, 2004, 16:16:35
Job time : 32 secs

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 16:16:06 ; Search time 195 Seconds
(without alignments)
415.589 Million cell updates/sec

Title: US-10-026-341A-2
Perfect score: 22
Sequence: 1 attcgatcggggcggggcgagc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		22	100.0	22	9	US-09-754-949-10	Sequence 10, Appl
2		22	100.0	22	9	US-09-918-889-20	Sequence 20, Appl
3		22	100.0	22	9	US-09-919-042-20	Sequence 20, Appl
4		22	100.0	22	13	US-10-417-422-11	Sequence 11, Appl
5		22	100.0	22	14	US-10-026-341A-2	Sequence 2, Appli
6		21	95.5	21	11	US-09-877-705A-81	Sequence 81, Appl
7	c	21	95.5	21	11	US-09-877-705A-82	Sequence 82, Appl
8		21	95.5	21	11	US-09-877-738A-81	Sequence 81, Appl
9	c	21	95.5	21	11	US-09-877-738A-82	Sequence 82, Appl
10		21	95.5	21	11	US-09-888-326-115	Sequence 115, App
11	c	21	95.5	21	11	US-09-888-326-229	Sequence 229, App
12		21	95.5	21	11	US-09-776-479-825	Sequence 825, App
13	c	21	95.5	21	11	US-09-776-479-826	Sequence 826, App
14		21	95.5	21	15	US-10-112-653-796	Sequence 796, App
15	c	21	95.5	21	15	US-10-112-653-797	Sequence 797, App
16		21	95.5	21	15	US-10-017-995-825	Sequence 825, App
17	c	21	95.5	21	15	US-10-017-995-826	Sequence 826, App
18	c	21	95.5	63	11	US-09-877-705A-149	Sequence 149, App
19	c	21	95.5	63	11	US-09-877-738A-149	Sequence 149, App
20		20.4	92.7	22	13	US-10-437-107-43	Sequence 43, Appl
21		20.4	92.7	22	15	US-10-052-092-43	Sequence 43, Appl
22		20	90.9	20	11	US-09-954-987B-104	Sequence 104, App
23		20	90.9	20	13	US-10-265-072-102	Sequence 102, App
24	c	17.4	79.1	183	15	US-10-145-289-10	Sequence 10, Appl
25		17.4	79.1	9646	13	US-10-311-455-1661	Sequence 1661, Ap
26	c	17.2	78.2	11115	13	US-09-769-734-49	Sequence 49, Appl
27	c	17.2	78.2	744802	12	US-10-292-798-1369	Sequence 1369, Ap
28		16.8	76.4	1073	13	US-10-027-632-248985	Sequence 248985,
29		16.8	76.4	1073	13	US-10-027-632-248986	Sequence 248986,
30		16.8	76.4	1073	14	US-10-027-632-248985	Sequence 248985,
31		16.8	76.4	1073	14	US-10-027-632-248986	Sequence 248986,
32		16.4	74.5	18	15	US-10-274-095-47	Sequence 47, Appl
33		16.4	74.5	6368	13	US-10-204-708-68	Sequence 68, Appl
34		16.4	74.5	6368	13	US-10-311-455-1912	Sequence 1912, Ap
35		16.4	74.5	6988	13	US-10-311-455-2414	Sequence 2414, Ap
36		16.4	74.5	10573	13	US-10-067-977-3	Sequence 3, Appli
37		16.4	74.5	63158	13	US-10-292-198-1	Sequence 1, Appli
38	c	16.2	73.6	498	11	US-09-918-995-11259	Sequence 11259, A
39		16.2	73.6	1599	15	US-10-156-761-1713	Sequence 1713, Ap
40		16.2	73.6	3586	10	US-09-769-207A-1	Sequence 1, Appli
41		16.2	73.6	3586	13	US-10-268-311-1	Sequence 1, Appli
42		16.2	73.6	6419	13	US-10-311-455-1331	Sequence 1331, Ap
43		16.2	73.6	6419	13	US-10-240-453-251	Sequence 251, App
44	c	16.2	73.6	9025608	15	US-10-156-761-1	Sequence 1, Appli
45	c	16	72.7	2263	15	US-10-190-471-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-754-949-10
; Sequence 10, Application US/09754949
; Patent No. US20020015939A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, JUSTIN
; APPLICANT: CORDELL, BARBARA
; TITLE OF INVENTION: METHODS FOR IDENTIFYING INHIBITORS OF
; TITLE OF INVENTION: NEURONAL DEGENERATION
; FILE REFERENCE: SCIOS.012A
; CURRENT APPLICATION NUMBER: US/09/754,949
; CURRENT FILING DATE: 2001-01-04
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-09-754-949-10

Query Match 100.0%; Score 22; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAGC 22
||||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAGC 22

RESULT 2
US-09-918-889-20
; Sequence 20, Application US/09918889
; Patent No. US20020053092A1
; GENERAL INFORMATION:
; APPLICANT: Readhead, Carol W.
; APPLICANT: Winston, Robert

```
;  APPLICANT:  Koeffler, H. Phillip
;  APPLICANT:  Muller, Carsten
;  TITLE OF INVENTION: NUCLEIC ACID CONSTRUCTS CONTAINING A
;  TITLE OF INVENTION:  CYCLIN A1 PROMOTER, AND KIT
;  FILE REFERENCE: 18810-81603
;  CURRENT APPLICATION NUMBER: US/09/918,889
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US 09/292,723
;  PRIOR FILING DATE: 1999-04-15
;  PRIOR APPLICATION NUMBER: US 09/191,920
;  PRIOR FILING DATE: 1998-11-13
;  PRIOR APPLICATION NUMBER: US 60/065,825
;  PRIOR FILING DATE: 1997-11-14
;  PRIOR APPLICATION NUMBER: PCT/US98/24238
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
;   LENGTH: 22
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Double-stranded oligonucleotide
US-09-918-889-20

  Query Match           100.0%;  Score 22;  DB 9;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 1;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 3
US-09-919-042-20
; Sequence 20, Application US/09919042
; Patent No. US20020056148A1
; GENERAL INFORMATION:
;  APPLICANT: Readhead, Carol W.
;  APPLICANT:  Winston, Robert
;  APPLICANT:  Koeffler, H. Phillip
;  APPLICANT:  Muller, Carsten
;  TITLE OF INVENTION: Transfection, Storage and Transfer of
;  TITLE OF INVENTION:  Male Germ Cells for Generation of Selectable Transgenic Stem
;  TITLE OF INVENTION:  Cells
;  FILE REFERENCE: 18810-81602
;  CURRENT APPLICATION NUMBER: US/09/919,042
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US 09/292,723
;  PRIOR FILING DATE: 1999-04-15
;  PRIOR APPLICATION NUMBER: US 09/191,920
;  PRIOR FILING DATE: 1998-11-13
;  PRIOR APPLICATION NUMBER: US 60/065,825
;  PRIOR FILING DATE: 1997-11-14
;  PRIOR APPLICATION NUMBER: PCT/US98/24238
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
;   LENGTH: 22
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Double-stranded oligonucleotide
US-09-919-042-20

  Query Match           100.0%;  Score 22;  DB 9;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 1;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 4
US-10-417-422-11
; Sequence 11, Application US/10417422
; Publication No. US20030219720A1
; GENERAL INFORMATION:
;  APPLICANT: MCCARTHY, JUSTIN
;  APPLICANT:  CORDELL, BARBARA
;  APPLICANT:  SCIOS, INC.
;  TITLE OF INVENTION: METHODS FOR IDENTIFYING INHIBITORS OF
;  TITLE OF INVENTION:  NEURONAL DEGENERATION
;  FILE REFERENCE: SCIOS.012C1
;  CURRENT APPLICATION NUMBER: US/10/417,422
;  CURRENT FILING DATE:  2003-04-14
;  PRIOR APPLICATION NUMBER: 09/754949
;  PRIOR FILING DATE: 2001-02-04
;  PRIOR APPLICATION NUMBER: 60/175200
;  PRIOR FILING DATE: 2000-01-10
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 22
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Oligonucleotide
US-10-417-422-11

  Query Match           100.0%;  Score 22;  DB 13;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 1;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22


RESULT 5
US-10-026-341A-2
; Sequence 2, Application US/10026341A
; Publication No. US20020137715A1
; GENERAL INFORMATION:
;  APPLICANT: Alain Mauviel
;  TITLE OF INVENTION: Blocking Sp1 Transcription Factor
;  TITLE OF INVENTION:  Broadly Inhibits Extracellular Matrix Gene Expression In
;  TITLE OF INVENTION:  Vitro and In Vivo: Implications for the Treatment of Tissue
;  TITLE OF INVENTION:  Fibrosis
;  FILE REFERENCE: MAU01.NP001
;  CURRENT APPLICATION NUMBER: US/10/026,341A
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/259,585
;  PRIOR FILING DATE: 2001-01-03
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 22
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic oligonucleotide
US-10-026-341A-2

  Query Match           100.0%;  Score 22;  DB 14;  Length 22;
  Best Local Similarity  100.0%;  Pred. No. 1;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAGC 22
```

```
RESULT 6
US-09-877-705A-81
; Sequence 81, Application US/09877705A
; Publication No. US20030008283A1
; GENERAL INFORMATION:
;  APPLICANT: Li, Jason
;  TITLE OF INVENTION: METHOD FOR SCREENING FOR DRUG CANDIDATES FOR MODULATING TRANSCRIP
;  TITLE OF INVENTION:  FACTOR ACTIVITY
;  FILE REFERENCE: 26757-704
;  CURRENT APPLICATION NUMBER: US/09/877,705A
;  CURRENT FILING DATE:  2001-08-16
;  NUMBER OF SEQ ID NOS: 162
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
;   LENGTH: 21
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Transcription factor probe PP81
US-09-877-705A-81

  Query Match             95.5%;  Score 21;  DB 11;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAG 21


RESULT 7
US-09-877-705A-82/c
; Sequence 82, Application US/09877705A
; Publication No. US20030008283A1
; GENERAL INFORMATION:
;  APPLICANT: Li, Jason
;  TITLE OF INVENTION: METHOD FOR SCREENING FOR DRUG CANDIDATES FOR MODULATING TRANSCRIP
;  TITLE OF INVENTION:  FACTOR ACTIVITY
;  FILE REFERENCE: 26757-704
;  CURRENT APPLICATION NUMBER: US/09/877,705A
;  CURRENT FILING DATE:  2001-08-16
;  NUMBER OF SEQ ID NOS: 162
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 82
;   LENGTH: 21
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Transcription factor probe PP82
US-09-877-705A-82

  Query Match             95.5%;  Score 21;  DB 11;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db         21 ATTCGATCGGGGCGGGGCGAG 1


RESULT 8
US-09-877-738A-81
; Sequence 81, Application US/09877738A
; Publication No. US20030022173A1
; GENERAL INFORMATION:
;  APPLICANT: Li, Jason
;  TITLE OF INVENTION: METHOD AND KIT FOR ISOLATING DNA PROBES THAT BIND TO ACTIVATED
;  TITLE OF INVENTION:  TRANSCRIPTION FACTORS
;  FILE REFERENCE: 26757-701
;  CURRENT APPLICATION NUMBER: US/09/877,738A
;  CURRENT FILING DATE:  2001-06-01
;  NUMBER OF SEQ ID NOS: 162
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
;   LENGTH: 21
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Transcription factor probe PP81
US-09-877-738A-81

  Query Match             95.5%;  Score 21;  DB 11;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db          1 ATTCGATCGGGGCGGGGCGAG 21


RESULT 9
US-09-877-738A-82/c
; Sequence 82, Application US/09877738A
; Publication No. US20030022173A1
; GENERAL INFORMATION:
;  APPLICANT: Li, Jason
;  TITLE OF INVENTION: METHOD AND KIT FOR ISOLATING DNA PROBES THAT BIND TO ACTIVATED
;  TITLE OF INVENTION:  TRANSCRIPTION FACTORS
;  FILE REFERENCE: 26757-701
;  CURRENT APPLICATION NUMBER: US/09/877,738A
;  CURRENT FILING DATE:  2001-06-01
;  NUMBER OF SEQ ID NOS: 162
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 82
;   LENGTH: 21
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Transcription factor probe PP82
US-09-877-738A-82

  Query Match             95.5%;  Score 21;  DB 11;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db         21 ATTCGATCGGGGCGGGGCGAG 1


RESULT 10
US-09-888-326-115
; Sequence 115, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, George
;  APPLICANT:  Hartmann, Gunther
;  TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
;  TITLE OF INVENTION:  Cell Lysis and Treating Cancer
;  FILE REFERENCE: C1039/7052 (AWS)
;  CURRENT APPLICATION NUMBER: US/09/888,326
;  CURRENT FILING DATE:  2001-06-22
;  PRIOR APPLICATION NUMBER: US 60/213,346
;  PRIOR FILING DATE: 2000-06-22
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 115
;   LENGTH: 21
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: phosphodiester backbone
US-09-888-326-115
```

Query Match 95.5%; Score 21; DB 11; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAG 21

RESULT 11
US-09-888-326-229/c
; Sequence 229, Application US/09888326
; Publication No. US20030026801A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, George
; APPLICANT: Hartmann, Gunther
; TITLE OF INVENTION: Methods for Enhancing Antibody-Induced
; TITLE OF INVENTION: Cell Lysis and Treating Cancer
; FILE REFERENCE: C1039/7052 (AWS)
; CURRENT APPLICATION NUMBER: US/09/888,326
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/213,346
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 848
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 229
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic oligonucleotide
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: phosphodiester backbone
US-09-888-326-229

Query Match 95.5%; Score 21; DB 11; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 21 ATTCGATCGGGGCGGGGCGAG 1

RESULT 12
US-09-776-479-825
; Sequence 825, Application US/09776479
; Publication No. US20030087848A1
; GENERAL INFORMATION:
; APPLICANT: Bratzler, Robert L.
; APPLICANT: Petersen, Deanna M.
; APPLICANT: Fouron, Yves
; TITLE OF INVENTION: Immunostimulatory Nucleic Acids for the
; TITLE OF INVENTION: Treatment of Asthma and Allergy
; FILE REFERENCE: C1037/7013 (HCL/MAT)
; CURRENT APPLICATION NUMBER: US/09/776,479
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 60/179,991
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 825
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Sequence
US-09-776-479-825

Query Match 95.5%; Score 21; DB 11; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAG 21

RESULT 13
US-09-776-479-826/c
; Sequence 826, Application US/09776479
; Publication No. US20030087848A1
; GENERAL INFORMATION:
; APPLICANT: Bratzler, Robert L.
; APPLICANT: Petersen, Deanna M.
; APPLICANT: Fouron, Yves
; TITLE OF INVENTION: Immunostimulatory Nucleic Acids for the
; TITLE OF INVENTION: Treatment of Asthma and Allergy
; FILE REFERENCE: C1037/7013 (HCL/MAT)
; CURRENT APPLICATION NUMBER: US/09/776,479
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 60/179,991
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 826
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Sequence
US-09-776-479-826

Query Match 95.5%; Score 21; DB 11; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 21 ATTCGATCGGGGCGGGGCGAG 1

RESULT 14
US-10-112-653-796
; Sequence 796, Application US/10112653
; Publication No. US20030050268A1
; GENERAL INFORMATION:
; APPLICANT: Krieg, Arthur M.
; APPLICANT: Berg, Daniel J.
; TITLE OF INVENTION: IMMUNOSTIMULATORY NUCLEIC ACID FOR
; TITLE OF INVENTION: TREATMENT OF NON-ALLERGIC INFLAMMATORY DISEASES
; FILE REFERENCE: C01039/70060(AWS)
; CURRENT APPLICATION NUMBER: US/10/112,653
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 60/279,642
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 1040
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 796
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-10-112-653-796

Query Match 95.5%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAG 21

RESULT 15
US-10-112-653-797/c
; Sequence 797, Application US/10112653
; Publication No. US20030050268A1
; GENERAL INFORMATION:
; APPLICANT: Krieg, Arthur M.
; APPLICANT: Berg, Daniel J.
; TITLE OF INVENTION: IMMUNOSTIMULATORY NUCLEIC ACID FOR
; TITLE OF INVENTION: TREATMENT OF NON-ALLERGIC INFLAMMATORY DISEASES
; FILE REFERENCE: C01039/70060(AWS)
; CURRENT APPLICATION NUMBER: US/10/112,653
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 60/279,642
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 1040
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 797
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-10-112-653-797

Query Match 95.5%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 21 ATTCGATCGGGGCGGGGCGAG 1

RESULT 16
US-10-017-995-825
; Sequence 825, Application US/10017995
; Publication No. US20030055014A1
; GENERAL INFORMATION:
; APPLICANT: Bratzler, Robert L.
; TITLE OF INVENTION: Inhibition of Angiogenesis by Nucleic Acids
; FILE REFERENCE: C1037/7025 (HCL/MAT)
; CURRENT APPLICATION NUMBER: US/10/017,995
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US 60/255,534
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 825
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Sequence
US-10-017-995-825

Query Match 95.5%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 1 ATTCGATCGGGGCGGGGCGAG 21

RESULT 17
US-10-017-995-826/c
; Sequence 826, Application US/10017995
; Publication No. US20030055014A1
; GENERAL INFORMATION:
; APPLICANT: Bratzler, Robert L.
; TITLE OF INVENTION: Inhibition of Angiogenesis by Nucleic Acids
; FILE REFERENCE: C1037/7025 (HCL/MAT)
; CURRENT APPLICATION NUMBER: US/10/017,995
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US 60/255,534
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 826
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Sequence
US-10-017-995-826

Query Match 95.5%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 21 ATTCGATCGGGGCGGGGCGAG 1

RESULT 18
US-09-877-705A-149/c
; Sequence 149, Application US/09877705A
; Publication No. US20030008283A1
; GENERAL INFORMATION:
; APPLICANT: Li, Jason
; TITLE OF INVENTION: METHOD FOR SCREENING FOR DRUG CANDIDATES FOR MODULATING TRANSCR
; TITLE OF INVENTION: FACTOR ACTIVITY
; FILE REFERENCE: 26757-704
; CURRENT APPLICATION NUMBER: US/09/877,705A
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 162
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 149
; LENGTH: 63
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hybridization probe MP82
US-09-877-705A-149

Query Match 95.5%; Score 21; DB 11; Length 63;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTCGATCGGGGCGGGGCGAG 21
||||||||||||||||||||
Db 63 ATTCGATCGGGGCGGGGCGAG 43

RESULT 19
US-09-877-738A-149/c
; Sequence 149, Application US/09877738A
; Publication No. US20030022173A1
; GENERAL INFORMATION:
; APPLICANT: Li, Jason
; TITLE OF INVENTION: METHOD AND KIT FOR ISOLATING DNA PROBES THAT BIND TO ACTIVATED
; TITLE OF INVENTION: TRANSCRIPTION FACTORS
; FILE REFERENCE: 26757-701
; CURRENT APPLICATION NUMBER: US/09/877,738A
; CURRENT FILING DATE: 2001-06-01
; NUMBER OF SEQ ID NOS: 162
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 149
; LENGTH: 63
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hybridization probe MP82
US-09-877-738A-149

```
  Query Match             95.5%;  Score 21;  DB 11;  Length 63;
  Best Local Similarity   100.0%;  Pred. No. 2.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||||||||||||||||
Db         63 ATTCGATCGGGGCGGGGCGAG 43


RESULT 20
US-10-437-107-43
; Sequence 43, Application US/10437107
; Publication No. US20030186313A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/437,107
;  CURRENT FILING DATE:  2003-05-13
;  PRIOR APPLICATION NUMBER: US/10/052,092
;  PRIOR FILING DATE: 2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
;   LENGTH: 22
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Estrogen Response Element
US-10-437-107-43

  Query Match             92.7%;  Score 20.4;  DB 13;  Length 22;
  Best Local Similarity   95.5%;  Pred. No. 5.5;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              |||||||| ||||||||||||
Db          1 ATTCGATCAGGGCGGGGCGAGC 22


RESULT 21
US-10-052-092-43
; Sequence 43, Application US/10052092
; Publication No. US20030027778A1
; GENERAL INFORMATION:
;  APPLICANT: Fuqua, Suzanne
;  APPLICANT:  Allred, D.
;  APPLICANT:  Hopp, Torsten A.
;  APPLICANT:  O'Connell, Peter
;  TITLE OF INVENTION: Methods and Composition in Breast Cancer Diagnosis and Therapeuti
;  FILE REFERENCE: P02102US2
;  CURRENT APPLICATION NUMBER: US/10/052,092
;  CURRENT FILING DATE:  2002-01-18
;  PRIOR APPLICATION NUMBER: US 60/262,990
;  PRIOR FILING DATE: 2001-01-19
;  PRIOR APPLICATION NUMBER: US 60/304,018
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43
;   LENGTH: 22
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Estrogen Response Element
US-10-052-092-43

  Query Match             92.7%;  Score 20.4;  DB 15;  Length 22;
  Best Local Similarity   95.5%;  Pred. No. 5.5;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              |||||||| ||||||||||||
Db          1 ATTCGATCAGGGCGGGGCGAGC 22


RESULT 22
US-09-954-987B-104
; Sequence 104, Application US/09954987B
; Publication No. US20030104523A1
; GENERAL INFORMATION:
;  APPLICANT: Stefan Bauer
;  APPLICANT:  Grayson B. Lipford
;  APPLICANT:  Hermann Wagner
;  TITLE OF INVENTION: PROCESS FOR HIGH THROUGHPUT SCREENING OF
;  TITLE OF INVENTION:  CpG-BASED IMMUNO-AGONIST/ANTAGONIST
;  FILE REFERENCE: C1041/7016 (AWS)
;  CURRENT APPLICATION NUMBER: US/09/954,987B
;  CURRENT FILING DATE:  2001-09-17
;  PRIOR APPLICATION NUMBER: US 60/233,035
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/263,657
;  PRIOR FILING DATE: 2001-01-23
;  PRIOR APPLICATION NUMBER: US 60/291,726
;  PRIOR FILING DATE: 2001-05-17
;  PRIOR APPLICATION NUMBER: US 60/300,210
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 230
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 104
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
US-09-954-987B-104

  Query Match             90.9%;  Score 20;  DB 11;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||
Db          1 TCGATCGGGGCGGGGCGAGC 20


RESULT 23
US-10-265-072-102
; Sequence 102, Application US/10265072
; Publication No. US20030166001A1
; GENERAL INFORMATION:
;  APPLICANT: Lipford, Grayson
;  TITLE OF INVENTION: TOLL-LIKE RECEPTOR 3 SIGNALING AGONISTS AND ANTAGONISTS
;  FILE REFERENCE: C01041.70031.US
;  CURRENT APPLICATION NUMBER: US/10/265,072
;  CURRENT FILING DATE:  2002-10-05
;  NUMBER OF SEQ ID NOS: 117
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 102
;   LENGTH: 20
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic oligonucleotide
US-10-265-072-102

  Query Match             90.9%;  Score 20;  DB 13;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
```

```
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||||
Db          1 TCGATCGGGGCGGGGCGAGC 20


RESULT 24
US-10-145-289-10/c
; Sequence 10, Application US/10145289
; Publication No. US20030077812A1
; GENERAL INFORMATION:
;  APPLICANT: James G. McArthur
;  APPLICANT:  Dale John Talbot
;  APPLICANT:  Andrew D. Simmons
;  APPLICANT:  Ryan McGuinness
;  APPLICANT:  Michael Kelly
;  APPLICANT:  Lisa V. Tsui
;  APPLICANT:  Thomas Dull
;  TITLE OF INVENTION: LENTIVIRAL VECTORS ENCODING CLOTTING
;  TITLE OF INVENTION:  FACTORS FOR GENE THERAPY
;  FILE REFERENCE: 131.2USU1
;  CURRENT APPLICATION NUMBER: US/10/145,289
;  CURRENT FILING DATE:  2002-05-14
;  PRIOR APPLICATION NUMBER: 60/291,083
;  PRIOR FILING DATE: 2001-05-14
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 183
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: enhancer element
US-10-145-289-10

  Query Match             79.1%;  Score 17.4;  DB 15;  Length 183;
  Best Local Similarity   94.7%;  Pred. No. 91;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||||||||||||||||||  |
Db        183 CGATCGGGGCGGGGCGAAC 165


RESULT 25
US-10-311-455-1661
; Sequence 1661, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1661
;   LENGTH: 9646
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1661

  Query Match             79.1%;  Score 17.4;  DB 13;  Length 9646;
  Best Local Similarity   94.7%;  Pred. No. 49;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
              ||||||||||||||||||
Db       5203 TCGATCGGGGCGGGGCGGG 5221


RESULT 26
US-09-769-734-49/c
; Sequence 49, Application US/09769734
; Publication No. US20030143666A1
; GENERAL INFORMATION:
;  APPLICANT: Ecopia BioSciences Inc.
;  TITLE OF INVENTION: Genetic Locus for Everninomicin Biosynthesis
;  FILE REFERENCE: PA 005-US
;  CURRENT APPLICATION NUMBER: US/09/769,734
;  CURRENT FILING DATE:  2001-01-26
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 49
;   LENGTH: 11115
;   TYPE: DNA
;   ORGANISM: M. carbonacea
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (8)..(1207)
;   OTHER INFORMATION: ORF 41 (positive strandedness)
;   OTHER INFORMATION: incomplete:  C-terminus only
;   NAME/KEY: misc_feature
;   LOCATION: (1213)..(2331)
;   OTHER INFORMATION: ORF 42 (positive strandedness)
;   NAME/KEY: misc_feature
;   LOCATION: (2364)..(3611)
;   OTHER INFORMATION: ORF 43 (positive strandedness)
;   NAME/KEY: misc_feature
;   LOCATION: (3623)..(4243)
;   OTHER INFORMATION: ORF 44 (positive strandedness)
;   NAME/KEY: misc_feature
;   LOCATION: (4149)..(5177)
;   OTHER INFORMATION: ORF 45  (positive strandedness)
;   NAME/KEY: misc_feature
;   LOCATION: (5177)..(6094)
;   OTHER INFORMATION: ORF 46 (negative strandedness)
;   NAME/KEY: misc_feature
;   LOCATION: (6271)..(7824)
;   OTHER INFORMATION: ORF 47 (negative strandedness)
;   NAME/KEY: misc_feature
;   LOCATION: (7903)..(8760)
;   OTHER INFORMATION: ORF 48 (negative strandedness)
;   NAME/KEY: misc_feature
;   LOCATION: (8781)..(9800)
;   OTHER INFORMATION: ORF 49  (negative strandedness)
US-09-769-734-49

  Query Match             78.2%;  Score 17.2;  DB 13;  Length 11115;
  Best Local Similarity   86.4%;  Pred. No. 59;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              |||||||||| |||||| | ||
Db       6118 ATTCGATCGGCGCGGGGGGGGC 6097


RESULT 27
US-10-292-798-1369/c
; Sequence 1369, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
```

```
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1369
;    LENGTH: 744802
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    LOCATION: source
;    FEATURE:
;    LOCATION: (1)..(744802)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (201)..(246)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (25640)..(25677)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (27078)..(27094)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (141192)..(141769)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (159571)..(159606)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (174525)..(174575)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (234891)..(235013)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (235514)..(235560)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (279677)..(279729)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (408660)..(409123)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (409204)..(409669)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (428381)..(428396)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (472204)..(472330)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (714252)..(714355)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (714447)..(714529)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (739794)..(739891)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (744484)..(744602)
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (51812)..(51911)
;    OTHER INFORMATION: a, t, c, g, unknown or other
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (57122)..(57221)
;    OTHER INFORMATION: a, t, c, g, unknown or other
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (79368)..(79467)
;    OTHER INFORMATION: a, t, c, g, unknown or other
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (293951)..(294050)
;    OTHER INFORMATION: a, t, c, g, unknown or other
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (310089)..(310188)
;    OTHER INFORMATION: a, t, c, g, unknown or other
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (332935)..(332935)
;    OTHER INFORMATION: a, t, c, g, unknown or other
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (332992)..(332992)
;    OTHER INFORMATION: a, t, c, g, unknown or other
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (362002)..(362101)
;    OTHER INFORMATION: a, t, c, g, unknown or other
;    FEATURE:
;    NAME/KEY: modified_base
;    LOCATION: (639781)..(639880)
;    OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-1369

  Query Match             78.2%;  Score 17.2;  DB 12;  Length 744802;
  Best Local Similarity   86.4%;  Pred. No. 31;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAGC 22
              ||||||| |||| |||||| ||
Db     384946 ATTCGATAGGGGTGGGGCGTGC 384925


RESULT 28
US-10-027-632-248985
; Sequence 248985, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 248985
;   LENGTH: 1073
```

```
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-248985

  Query Match             76.4%;  Score 16.8;  DB 13;  Length 1073;
  Best Local Similarity   90.0%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAG 21
              ||| |||||||||||| |||
Db        922 TTCAATCGGGGCGGGGGGAG 941


RESULT 29
US-10-027-632-248986
; Sequence 248986, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 248986
;   LENGTH: 1073
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-248986

  Query Match             76.4%;  Score 16.8;  DB 13;  Length 1073;
  Best Local Similarity   90.0%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAG 21
              ||| |||||||||||| |||
Db        922 TTCAATCGGGGCGGGGGGAG 941


RESULT 30
US-10-027-632-248985
; Sequence 248985, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 248985
;   LENGTH: 1073
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-248985

  Query Match             76.4%;  Score 16.8;  DB 14;  Length 1073;
  Best Local Similarity   90.0%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAG 21
              ||| |||||||||||| |||
Db        922 TTCAATCGGGGCGGGGGGAG 941


RESULT 31
US-10-027-632-248986
; Sequence 248986, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 248986
;   LENGTH: 1073
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-248986

  Query Match             76.4%;  Score 16.8;  DB 14;  Length 1073;
  Best Local Similarity   90.0%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAG 21
              ||| |||||||||||| |||
Db        922 TTCAATCGGGGCGGGGGGAG 941


RESULT 32
US-10-274-095-47
; Sequence 47, Application US/10274095
; Publication No. US20030120433A1
; GENERAL INFORMATION:
;  APPLICANT: Yokota, Hiroki
;  APPLICANT:  Sun, Hui Bin
;  TITLE OF INVENTION: Methods for Predicting Transcription
;  TITLE OF INVENTION:  Levels
```

; FILE REFERENCE: ARTI.0137US
; CURRENT APPLICATION NUMBER: US/10/274,095
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: 60/329,961
; PRIOR FILING DATE: 2001-10-17
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 47
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fragment
US-10-274-095-47

Query Match 74.5%; Score 16.4; DB 15; Length 18;
Best Local Similarity 94.4%; Pred. No. 3.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GATCGGGGCGGGGCGAGC 22
||||||||||||||||| |
Db 1 GATCGGGGCGGGGCGATC 18

RESULT 33
US-10-204-708-68
; Sequence 68, Application US/10204708
; Publication No. US20030141852A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 68
; LENGTH: 6368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-68

Query Match 74.5%; Score 16.4; DB 13; Length 6368;
Best Local Similarity 94.4%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GATCGGGGCGGGGCGAGC 22
||||||||||||||||| ||
Db 4207 GATCGGGGCGGGGCGGGC 4224

RESULT 34
US-10-311-455-1912
; Sequence 1912, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Dete
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1912
; LENGTH: 6368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1912

Query Match 74.5%; Score 16.4; DB 13; Length 6368;
Best Local Similarity 94.4%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GATCGGGGCGGGGCGAGC 22
||||||||||||||||| ||
Db 4207 GATCGGGGCGGGGCGGGC 4224

RESULT 35
US-10-311-455-2414
; Sequence 2414, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Dete
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2414
; LENGTH: 6988
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2414

Query Match 74.5%; Score 16.4; DB 13; Length 6988;
Best Local Similarity 94.4%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAG 21
||| ||||||||||||||
Db 1297 CGAGCGGGGCGGGGCGAG 1314

RESULT 36
US-10-067-977-3
; Sequence 3, Application US/10067977
; Publication No. US20030157679A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua and KE, Zhaoxi
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

```
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001313
;  CURRENT APPLICATION NUMBER: US/10/067,977
;  CURRENT FILING DATE:  2002-02-08
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 10573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-067-977-3

  Query Match             74.5%;  Score 16.4;  DB 13;  Length 10573;
  Best Local Similarity   94.4%;  Pred. No. 1.4e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAG 21
              || ||||||||||||||||
Db       2862 CGGTCGGGGCGGGGCGAG 2879


RESULT 37
US-10-292-198-1
; Sequence 1, Application US/10292198
; Publication No. US20030157654A1
; GENERAL INFORMATION:
;  APPLICANT: SHEN, Ben
;  APPLICANT:  LIU, Wen
;  TITLE OF INVENTION: BIOSYTHESIS OF ENEDIYNE COMPOUNDS BY MANIPULATION OF C-1027 GENE
;  TITLE OF INVENTION:  PATHWAY
;  FILE REFERENCE: 054030-0007
;  CURRENT APPLICATION NUMBER: US/10/292,198
;  CURRENT FILING DATE:  2003-03-14
;  PRIOR APPLICATION NUMBER: US 10/159,257
;  PRIOR FILING DATE: 2002-05-31
;  PRIOR APPLICATION NUMBER: US 09/478,188
;  PRIOR FILING DATE: 2000-01-05
;  PRIOR APPLICATION NUMBER: US 60/115,434
;  PRIOR FILING DATE: 1999-01-06
;  NUMBER OF SEQ ID NOS: 146
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 63158
;   TYPE: DNA
;   ORGANISM: Streptomyces globisporus
US-10-292-198-1

  Query Match             74.5%;  Score 16.4;  DB 13;  Length 63158;
  Best Local Similarity   94.4%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          5 GATCGGGGCGGGGCGAGC 22
              || ||||||||||||||||
Db      46469 GAGCGGGGCGGGGCGAGC 46486


RESULT 38
US-09-918-995-11259/c
; Sequence 11259, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11259
;   LENGTH: 498
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(498)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-11259

  Query Match             73.6%;  Score 16.2;  DB 11;  Length 498;
  Best Local Similarity   85.7%;  Pred. No. 2.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAGC 22
              ||| || ||||||||||| ||
Db         26 TTCTATTGGGGCGGGGCGTGC 6


RESULT 39
US-10-156-761-1713
; Sequence 1713, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1713
;   LENGTH: 1599
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1599)
US-10-156-761-1713

  Query Match             73.6%;  Score 16.2;  DB 15;  Length 1599;
  Best Local Similarity   85.7%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAGC 22
              ||||| ||||||| |||| ||
Db        607 TTCGAACGGGGCGTGGCGGGC 627


RESULT 40
US-09-769-207A-1
; Sequence 1, Application US/09769207A
; Patent No. US20020132234A1
; GENERAL INFORMATION:
;  APPLICANT: DzGenes, LLC
;  TITLE OF INVENTION: NITRIC OXIDE SYNTHASE GENE DIAGNOSTIC POLYMORPHISMS
;  FILE REFERENCE: DZG 2165.1
;  CURRENT APPLICATION NUMBER: US/09/769,207A
;  CURRENT FILING DATE:  2001-01-24
;  PRIOR APPLICATION NUMBER: US 60/177,775
;  PRIOR FILING DATE: 2000-01-24
;  PRIOR APPLICATION NUMBER: US 60/220,662
;  PRIOR FILING DATE: 2000-07-25
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.0
```

; SEQ ID NO 1
; LENGTH: 3586
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1465)..(3585)
; OTHER INFORMATION: Promotor region and exon 1, partial CDS
; NAME/KEY: mRNA
; LOCATION: (3473)..(3585)
; NAME/KEY: exon
; LOCATION: (3473)..(3586)
; NAME/KEY: gene
; LOCATION: (3473)..(3585)
; OTHER INFORMATION: gene=NOS3
; NAME/KEY: CDS
; LOCATION: (3494)..(3586)
; OTHER INFORMATION: n=unknown
US-09-769-207A-1

Query Match 73.6%; Score 16.2; DB 10; Length 3586;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              ||  ||| ||||||||||||||
Db       3360 ATGGGATAGGGGCGGGGCGAG 3380
```

Search completed: February 18, 2004, 17:16:29
Job time : 201 secs

OM nucleic - nucleic search, using sw model

Run on: February 18, 2004, 16:10:46 ; Search time 1633 Seconds
(without alignments)
327.433 Million cell updates/sec

Title: US-10-026-341A-2
Perfect score: 22
Sequence: 1 attcgatcggggcggggcgagc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	20	90.9	429	9	AT005447	AT005447 AT005447
	2	19	86.4	311	10	BE654158	BE654158 UI-M-AN1-
c	3	18.4	83.6	1640	12	BM805163	BM805163 AGENCOURT
	4	17.8	80.9	490	28	BH629785	BH629785 1007082B1
c	5	17.8	80.9	495	14	CA708962	CA708962 wdk2c.pk0
c	6	17.8	80.9	552	29	CC042414	CC042414 3591_1_15
c	7	17.8	80.9	603	14	CD228564	CD228564 CCC1_8_H0
	8	17.8	80.9	710	29	BZ997350	BZ997350 PUGGL63TB
	9	17.8	80.9	723	29	BZ339055	BZ339055 ic29d10.g
c	10	17.8	80.9	896	14	CD439046	CD439046 EL01N0520
	11	17.8	80.9	1511	12	BM552949	BM552949 AGENCOURT
c	12	17.8	80.9	1557	11	AY106281	AY106281 Zea mays
	13	17.4	79.1	319	14	CD550345	CD550345 B0310F12-
	14	17.4	79.1	322	14	CD550569	CD550569 B0315A09-
	15	17.4	79.1	341	13	BY204565	BY204565 BY204565
	16	17.4	79.1	350	13	BY302682	BY302682 BY302682
	17	17.4	79.1	355	13	BY210660	BY210660 BY210660
c	18	17.4	79.1	375	9	AW122220	AW122220 UI-M-BH2.
c	19	17.4	79.1	379	9	AI846527	AI846527 UI-M-AN1-
	20	17.4	79.1	406	9	AA615769	AA615769 vo72f08.r
	21	17.4	79.1	418	14	CB765932	CB765932 AMGNNUC:M
	22	17.4	79.1	423	12	BI220387	BI220387 602935686
	23	17.4	79.1	437	10	BB859799	BB859799 BB859799
	24	17.4	79.1	448	9	AI194235	AI194235 ue52e11.r
c	25	17.4	79.1	472	14	CB640552	CB640552 OSJNEa15E
	26	17.4	79.1	493	10	BE650594	BE650594 UI-M-BH2.
	27	17.4	79.1	523	4	BX529827	Bx529827 RZPD Mus
	28	17.4	79.1	530	10	BF015640	BF015640 uy27b02.y
	29	17.4	79.1	600	13	BU919508	BU919508 6024-75 M
	30	17.4	79.1	600	13	BU924221	BU924221 7082-67 M
	31	17.4	79.1	624	14	CB578330	CB578330 AMGNNUC:N
	32	17.4	79.1	635	14	BY721913	BY721913 BY721913
	33	17.4	79.1	645	14	BY721911	BY721911 BY721911
	34	17.4	79.1	679	12	BI111370	BI111370 602899252
	35	17.4	79.1	704	29	BZ658953	BZ658953 OGAMF20TM
	36	17.4	79.1	709	13	BQ445767	BQ445767 UI-M-ER0-
	37	17.4	79.1	735	13	BQ444735	BQ444735 UI-M-ER0-
	38	17.4	79.1	741	13	BQ445216	BQ445216 UI-M-ER0-
	39	17.4	79.1	764	12	BI146925	BI146925 602911551
	40	17.4	79.1	773	14	BY712338	BY712338 BY712338
	41	17.4	79.1	774	10	BG172820	BG172820 602337255
	42	17.4	79.1	776	13	BQ746273	BQ746273 UI-M-ER0-
c	43	17.4	79.1	781	14	CB657329	CB657329 OSJNEc12J
	44	17.4	79.1	791	12	BI100531	BI100531 602886667
	45	17.4	79.1	793	13	BU525924	BU525924 AGENCOURT

ALIGNMENTS

RESULT 1
AT005447/c
LOCUS AT005447 429 bp mRNA linear EST 25-MAR-2002
DEFINITION AT005447 POMFBO1 Pleurotus ostreatus cDNA clone MFB34-F01, mRNA sequence.
ACCESSION AT005447
VERSION AT005447.1 GI:13420306
KEYWORDS EST.
SOURCE Pleurotus ostreatus (oyster mushroom)
ORGANISM Pleurotus ostreatus
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Pleurotaceae; Pleurotus.
REFERENCE 1 (bases 1 to 429)
AUTHORS Lee,S.H., Kim,B.G., Kim.K.J., Lee,J.S., Yun,D.W., Hahn,J.H., Kim ,G.H., Lee,K.H., Suh,D.S., Kwon,S.T., Lee,C.S. and Yoo,Y.B.
TITLE Comparative Analysis of Sequences Expressed during the Liquid-Cultured Mycelia and Fruit Body Stages of Pleurotus ostreatus
JOURNAL Fungal Genet. Biol. 35 (2), 115-134 (2002)
MEDLINE 21838665
PUBMED 11848675
COMMENT Contact: Beom-Gi Kim
Division of applied microbiology
Institute of Agricultural Science and Technology(NIAST)
249 Seodundong Kweonseonku, Suwon 441707, Korea
Tel: 82-331-290-0347
Fax: 82-331-290-0399

```
            Email: bgkimyes@rda.go.kr
            Submitted through BRIC(Biological Research Information Center) of
            Korea
            URL: http://bric.postech.ac.kr/
            GeneNuri No. KS105130.
FEATURES              Location/Qualifiers
     source           1. .429
                      /organism="Pleurotus ostreatus"
                      /mol_type="mRNA"
                      /cultivar="ASI 2029"
                      /db_xref="taxon:5322"
                      /clone="MFB34-F01"
                      /dev_stage="mature fruiting body"
                      /lab_host="E.coli"
                      /clone_lib="POMFBO1"
                      /note="Vector: lambda TripEX2; Site_1: SfiIA; Site_2:
                      SfiIBI; average insert size:1500 bp; initial pfu:5 * 10 7;
                      Isolation of total RNA from the mature fruiting body
                      cultivated in poplar tree sawdust bottle"
BASE COUNT      101 a    135 c     92 g    101 t
ORIGIN

  Query Match             90.9%;  Score 20;  DB 9;  Length 429;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAG 21
              ||||||||||||||||||||
Db        120 TTCGATCGGGGCGGGGCGAG 101


RESULT 2
BE654158
LOCUS       BE654158                311 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  UI-M-AN1-aff-f-04-0-UI.r2 NIH_BMAP_MBG_N Mus musculus cDNA clone
            UI-M-AN1-aff-f-04-0-UI 5', mRNA sequence.
ACCESSION   BE654158
VERSION     BE654158.1  GI:9980071
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements The following repetitive elements were
            found in this cDNA sequence: 119-174, >(CA)n#Simple_repeat
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .311
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-AN1-aff-f-04-0-UI"
                      /dev_stage="27-32 days"
                      /lab_host="DH10B (Life Technologies)"
                      /clone_lib="NIH_BMAP_MBG_N"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_MBG_N library is a normalized library constructed
                      from  mouse basal ganglia. The tag is a string of 5
                      nucleotides present  between the Not I site and the
                      oligo-dT track.  The library was  constructed as described
                      by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                      , 1996. Tissue provided by Ms. Annie Novakovich,
                      Zivic-Miller Laboratories."
BASE COUNT       62 a     62 c    116 g     71 t
ORIGIN

  Query Match             86.4%;  Score 19;  DB 10;  Length 311;
  Best Local Similarity   100.0%;  Pred. No. 7.5e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              |||||||||||||||||||
Db        217 CGATCGGGGCGGGGCGAGC 235


RESULT 3
BM805163/c
LOCUS       BM805163                1640 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6490256 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5726046
            5', mRNA sequence.
ACCESSION   BM805163
VERSION     BM805163.1  GI:19121986
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1640)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12717  row: o  column: 07
            High quality sequence start: 17
            High quality sequence stop: 440.
FEATURES              Location/Qualifiers
     source           1. .1640
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:5726046"
                      /lab_host="DH10B"
                      /clone_lib="NIH_MGC_125"
                      /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                      Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                      of three ovaries, from females ranging in age from 38 to
                      49 yo. Library is oligo-dT primed and directionally cloned
                      (EcoRV site is destroyed upon cloning). Average insert
                      size 2.1 kb, insert size range 1-3.5 kb. Library is
                      normalized and enriched for full-length clones and was
                      constructed by C. Gruber (Invitrogen). Research Genetics
                      tracking code 036."
BASE COUNT      222 a    608 c    438 g    369 t      3 others
ORIGIN

  Query Match             83.6%;  Score 18.4;  DB 12;  Length 1640;
  Best Local Similarity   95.0%;  Pred. No. 1e+03;
```

```
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGA 20
              || ||||||||||||||||||
Db       1456 ATGCGATCGGGGCGGGGCGA 1437


RESULT 4
BH629785
LOCUS       BH629785                490 bp    DNA     linear   GSS 30-JAN-2002
DEFINITION  1007082B12.2EL_y1 1007 - RescueMu Grid H Zea mays genomic, genomic
            survey sequence.
ACCESSION   BH629785
VERSION     BH629785.1  GI:18443036
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1007082  column: 31
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="1007 - RescueMu Grid H"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units.  For more information on RescueMu, go to the  web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.'  Grid H was grown at Berkeley in 2001.  DNA
                     was extracted from leaf punches,  double digested using
                     BamHI and BglII, and ligated to form circular plasmids.
                     DH10B cells were transformed and then screened on LB
                     plates with ampicillin."
BASE COUNT       76 a    153 c    196 g     65 t
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 28;  Length 490;
  Best Local Similarity   90.5%;  Pred. No. 2.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAGC 22
              ||||||||| ||| |||||||
Db        345 TTCGATCGGAGCGCGGCGAGC 365


RESULT 5
CA708962/c
LOCUS       CA708962                495 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  wdk2c.pk011.a20 wdk2c Triticum aestivum cDNA clone wdk2c.pk011.a20
            5' end, mRNA sequence.
ACCESSION   CA708962
VERSION     CA708962.1  GI:25430755
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
            Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE     DuPont Wheat cDNA Sequence
  JOURNAL   Unpublished
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone="wdk2c.pk011.a20"
                     /tissue_type="kernel"
                     /clone_lib="wdk2c"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
                     days after anthesis."
BASE COUNT       91 a    165 c    165 g     71 t      3 others
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 14;  Length 495;
  Best Local Similarity   90.5%;  Pred. No. 2.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||| |||| |||||||||
Db         76 ATTCGACCGGGCCGGGGCGAG 56


RESULT 6
CC042414/c
LOCUS       CC042414                552 bp    DNA     linear   GSS 01-APR-2003
DEFINITION  3591_1_151_1_D02.y_1 3591 - RescueMu Grid P Zea mays genomic,
            genomic survey sequence.
ACCESSION   CC042414
VERSION     CC042414.1  GI:29457305
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 3591_1_151_1  row: 28
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .552
```

```
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73/K55"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="3591 - RescueMu Grid P"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid P was grown at Molokai in 2002. DNA was
                     extracted from leaf strips, double digested using BamHI
                     and BglII, and ligated to form circular plasmids. DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
BASE COUNT       75 a    203 c    176 g     98 t
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 29;  Length 552;
  Best Local Similarity   90.5%;  Pred. No. 2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAGC 22
              |||||||||| ||| |||||||
Db        105 TTCGATCGGAGCGCGGCGAGC 85


RESULT 7
CD228564/c
LOCUS       CD228564                603 bp    mRNA    linear   EST 21-MAY-2003
DEFINITION  CCC1_8_H03.g1_A007 Callus culture/cell suspension Sorghum bicolor
            cDNA clone CCC1_8_H03_A007 5', mRNA sequence.
ACCESSION   CD228564
VERSION     CD228564.1  GI:30971998
KEYWORDS    EST.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S., Klein
            ,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Rathore,K., Eastman
            ,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: callus culture and cell suspension
  JOURNAL   Unpublished
COMMENT     Other_ESTs: CCC1_8_H03.b1_A007
            Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
            the Human Genome Center, University of Tokyo Institute of Medical
            Science; plant material and RNA prepared at Texas A & M University;
            sequencing done in the Laboratory for Genomics and Bioinformatics,
            University of Georgia. Sequence ends have been trimmed to exclude
            vector and regions below Phred quality 16. Three-prime sequences
            are presented as their reverse complement and have been trimmed to
            exclude polyA.
            Seq primer: Sug5 (CTTCTGCTCTAAAAGCTGCG).
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Sorghum bicolor"
                     /mol_type="mRNA"
                     /cultivar="RTx430"
                     /db_xref="taxon:4558"
                     /clone="CCC1_8_H03_A007"
                     /lab_host="DH10B-T1 phage-resistant E. coli"
                     /clone_lib="Callus culture/cell suspension"
                     /note="Vector: pME18S-FL3; Site_1: XhoI; Site_2: XhoI; The
                     library was prepared from a mixture of polyA+ RNA from
                     callus culture tissue and cells in suspension culture.
                     Double-stranded cDNA was cloned unidirectionally into
                     different DraIII sites of the pME18S-FL3 vector (5-prime
                     DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG
                     ). XhoI excises the cDNA insert."
BASE COUNT       96 a    229 c    189 g     89 t
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 14;  Length 603;
  Best Local Similarity   90.5%;  Pred. No. 2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAGC 22
              |||||||||| ||| |||||||
Db        234 TTCGATCGGAGCGCGGCGAGC 214


RESULT 8
BZ997350
LOCUS       BZ997350                710 bp    DNA     linear   GSS 25-MAR-2003
DEFINITION  PUGGL63TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa378L06,
            genomic survey sequence.
ACCESSION   BZ997350
VERSION     BZ997350.1  GI:29240767
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Other_GSSs: PUGGL63TD
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .710
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa378L06"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      162 a    194 c    205 g    149 t
ORIGIN

  Query Match             80.9%;  Score 17.8;  DB 29;  Length 710;
  Best Local Similarity   90.5%;  Pred. No. 2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATTCGATCGGGGCGGGGCGAG 21
              |||||||| || |||||||||
Db        228 ATTCGATCCGGCCGGGGCGAG 248


RESULT 9
```

```
BZ339055
LOCUS       BZ339055                723 bp    DNA     linear   GSS 06-NOV-2002
DEFINITION  ic29d10.g1 WGS-SbicolorF (JM107 adapted methyl filtered) Sorghum
            bicolor genomic clone ic29d10 5', genomic survey sequence.
ACCESSION   BZ339055
VERSION     BZ339055.1  GI:24735457
KEYWORDS    GSS.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., Palmer,L., McCombie,W.R. and Martienssen,R.A.
  TITLE     Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
  JOURNAL   Unpublished
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: ic29  row: d  column: 10
            Seq primer: -21M13UnivRev
            Class: shotgun
            High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4558"
                     /clone="ic29d10"
                     /lab_host="JM107 or DH5a"
                     /clone_lib="WGS-SbicolorF (JM107 adapted methyl filtered)"
                     /note="Site_1: Xba I; Site_2: Xba I; The vector was
                     digested with XbaI and one nucleotide was added by fill in
                     in the recessive 3' end. The genomic DNA was nebulized,
                     end repaired, adaptor ligated and size fractionated using
                     sephadex. The resulting fragments were between 0.8 and 3
                     kb and were cloned into the vector (.x/y reads in M13mp19,
                     .b/g reads in pUC19). The same ligation was transformed in
                     either JM107 or DH5a."
BASE COUNT      153 a    189 c    217 g    163 t      1 others
ORIGIN

  Query Match            80.9%;  Score 17.8;  DB 29;  Length 723;
  Best Local Similarity  90.5%;  Pred. No. 2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAGC 22
              ||||||||| ||| |||||||
Db         78 TTCGATCGGAGCGCGGCGAGC 98


RESULT 10
CD439046/c
LOCUS       CD439046                896 bp    mRNA    linear   EST 03-JUN-2003
DEFINITION  EL01N0520C10.b Endosperm_5 Zea mays cDNA, mRNA sequence.
ACCESSION   CD439046
VERSION     CD439046.1  GI:31354689
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 896)
  AUTHORS   Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
            Messing,J.
  TITLE     Sequencing of the maize endosperm ESTs
  JOURNAL   Unpublished
COMMENT     Contact: Lai, Jinsheng
            Dr. Joachim Messing's lab
            Waksman Institute, Rutgers University
            190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
            Tel: 732-445-3801
            Fax: 732-445-5735
            Email: jlai@waksman.rutgers.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .896
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="W22"
                     /db_xref="taxon:4577"
                     /tissue_type="Endosperm of 7-23DAP"
                     /clone_lib="Endosperm_5"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      212 a    255 c    238 g    191 t
ORIGIN

  Query Match            80.9%;  Score 17.8;  DB 14;  Length 896;
  Best Local Similarity  90.5%;  Pred. No. 1.9e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 TTCGATCGGGGCGGGGCGAGC 22
              ||||||||| ||| |||||||
Db        113 TTCGATCGGAGCGCGGCGAGC 93


RESULT 11
BM552949
LOCUS       BM552949                1511 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6558391 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742743
            5', mRNA sequence.
ACCESSION   BM552949
VERSION     BM552949.1  GI:18791249
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1511)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12761  row: f  column: 24
            High quality sequence start: 178
            High quality sequence stop: 340.
FEATURES             Location/Qualifiers
     source          1. .1511
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5742743"
                     /tissue_type="medulla"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_119"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source normal medulla from
                     anonymous male age 27.  Library  is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.3 kb, insert size range
```

0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."
BASE COUNT 428 a 413 c 450 g 215 t 5 others
ORIGIN

Query Match 80.9%; Score 17.8; DB 12; Length 1511;
Best Local Similarity 90.5%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTCGATCGGGGCGGGGCGAGC 22
||||| |||||||||| ||||
Db 54 TTCGACCGGGGCGGGGGGAGC 74

RESULT 12
AY106281/c
LOCUS AY106281 1557 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO117884 mRNA sequence.
ACCESSION AY106281
VERSION AY106281.1 GI:21209359
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1557)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1557)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES Location/Qualifiers
source 1. .1557
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:636873"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT 424 a 382 c 377 g 374 t
ORIGIN

Query Match 80.9%; Score 17.8; DB 11; Length 1557;
Best Local Similarity 90.5%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTCGATCGGGGCGGGGCGAGC 22
||||||||| ||| |||||||
Db 231 TTCGATCGGAGCGCGGCGAGC 211

RESULT 13
CD550345
LOCUS CD550345 319 bp mRNA linear EST 11-JUN-2003
DEFINITION B0310F12-5 NIA Mouse E9.5 Whole Embryo cDNA Library (Long) Mus musculus cDNA clone NIA:B0310F12 IMAGE:30430055 5', mRNA sequence.
ACCESSION CD550345
VERSION CD550345.1 GI:31598076
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 319)
AUTHORS Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE 21429098
PUBMED 11544199
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: B0310 row: F column: 12
Seq primer: M13 Reverse
High quality sequence stop: 319
POLYA=No.
FEATURES Location/Qualifiers
source 1. .319
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:B0310F12-5"
/db_xref="taxon:10090"
/clone="NIA:B0310F12 IMAGE:30430055"
/tissue_type="E9.5 whole embryo"
/dev_stage="whole embryo including extraembryonic tissues at 9.5-days postcoitum"
/lab_host="DH10B"
/clone_lib="NIA Mouse E9.5 Whole Embryo cDNA Library (Long)"
/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]).Total RNAs were extracted from a pool of 16 embryos at 9.5-days postcoitum. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT-3'] from6.1 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0kb. The library was constructed by Yulan Piao."
BASE COUNT 73 a 87 c 108 g 51 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 14; Length 319;
Best Local Similarity 94.7%; Pred. No. 3.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22

```
                    ||| ||||||||||||||||
Db          60 CGAGCGGGGCGGGGCGAGC 78


RESULT 14
CD550569
LOCUS       CD550569                322 bp    mRNA    linear   EST 11-JUN-2003
DEFINITION  B0315A09-5 NIA Mouse E9.5 Whole Embryo cDNA Library (Long) Mus
            musculus cDNA clone NIA:B0315A09 IMAGE:30430472 5', mRNA sequence.
ACCESSION   CD550569
VERSION     CD550569.1  GI:31598300
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 322)
  AUTHORS   Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
  TITLE     Construction of long-transcript enriched cDNA libraries from
            submicrogram amounts of total RNAs by a universal PCR amplification
            method
  JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
  MEDLINE   21429098
   PUBMED   11544199
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: B0315  row: A  column: 09
            Seq primer: M13 Reverse
            High quality sequence stop: 322
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .322
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:B0315A09-5"
                     /db_xref="taxon:10090"
                     /clone="NIA:B0315A09 IMAGE:30430472"
                     /tissue_type="E9.5 whole embryo"
                     /dev_stage="whole embryo including extraembryonic tissues
                     at 9.5-days postcoitum"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse E9.5 Whole Embryo cDNA Library (Long
                     )"
                     /note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
                     Site_2: NotI; Mouse cDNA project by the Laboratory of
                     Genetics, National Institute on Aging (NIA), Intramural
                     Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
                     This is a long-transcript enriched cDNA library (Ref.
                     Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]).Total
                     RNAs were extracted from a pool of 16 embryos at 9.5-days
                     postcoitum. Double-stranded cDNAs were synthesized with an
                     Oligo(dT) primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3']
                     from6.1 ug of total RNA, treated with T4 DNA polymerase,
                     and purified by ethanol-precipitation. The cDNAs were
                     ligated to Lone-linker LL-Sal4, purified by
                     phenol/chloroform, and separated from free linkers by
                     Centricon 100. Then, the cDNAs were amplified by
                     long-range high fidelity PCR using Ex Taq polymerase
                     (Takara) with a primer Sal4-S. The products were purified
                     by phenol/chloroform and Centricon 100. The cDNAs were
                     digested with SalI and NotI enzymes and cloned into
                     SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E.
                     coli host was transformed with the ligation mixture by the
                     standard chemical method. The average insert size is about
                     3.0kb. The library was constructed by Yulan Piao."
BASE COUNT       74 a     88 c    109 g     51 t
ORIGIN
```

```
  Query Match           79.1%;  Score 17.4;  DB 14;  Length 322;
  Best Local Similarity  94.7%;  Pred. No. 3.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db         60 CGAGCGGGGCGGGGCGAGC 78


RESULT 15
BY204565
LOCUS       BY204565                341 bp    mRNA    linear   EST 10-DEC-2002
DEFINITION  BY204565 RIKEN full-length enriched, B6-derived CD11 +ve dendritic
            cells Mus musculus cDNA clone F730213P11 5', mRNA sequence.
ACCESSION   BY204565
VERSION     BY204565.1  GI:26384371
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
```

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .341
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="F730213P11"
                     /cell_type="B6-derived CD11 +ve dendritic cells"
                     /clone_lib="RIKEN full-length enriched, B6-derived CD11
                     +ve dendritic cells"
BASE COUNT       64 a     77 c    130 g     70 t
ORIGIN
```

Query Match 79.1%; Score 17.4; DB 13; Length 341;
Best Local Similarity 94.7%; Pred. No. 3.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        136 CGAGCGGGGCGGGGCGAGC 154
```

RESULT 16
BY302682

```
LOCUS       BY302682                350 bp    mRNA    linear   EST 11-DEC-2002
DEFINITION  BY302682 RIKEN full-length enriched, 14.5 days embryo df/df
            Rathke's pouches Mus musculus cDNA clone K820005D04 5', mRNA
            sequence.
ACCESSION   BY302682
VERSION     BY302682.1  GI:26493019
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 350)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Michelle Brinkmeier and Sally Camper (
            Dept. Human Genetics University of Michigan Medical School 4301
            MSRB 3 1500 W. Medical Center Dr. Ann Arbor,MI 48109-0638 USA )
            whose assistance we gratefully acknowledge.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .350
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="K820005D04"
                     /tissue_type="Rathke's pouches"
                     /dev_stage="14.5 days embryo df/df"
                     /clone_lib="RIKEN full-length enriched, 14.5 days embryo
                     df/df Rathke's pouches"
BASE COUNT       77 a     84 c    127 g     62 t
ORIGIN
```

Query Match 79.1%; Score 17.4; DB 13; Length 350;
Best Local Similarity 94.7%; Pred. No. 3.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        111 CGAGCGGGGCGGGGCGAGC 129
```

RESULT 17
BY210660

```
LOCUS       BY210660                355 bp    mRNA    linear   EST 10-DEC-2002
```

DEFINITION BY210660 RIKEN full-length enriched, B6-derived CD11 +ve dendritic cells Mus musculus cDNA clone F730317H16 5', mRNA sequence.
ACCESSION BY210660
VERSION BY210660.1 GI:26391233
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 355)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .355
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F730317H16"
/cell_type="B6-derived CD11 +ve dendritic cells"
/clone_lib="RIKEN full-length enriched, B6-derived CD11 +ve dendritic cells"
BASE COUNT 70 a 79 c 133 g 71 t 2 others
ORIGIN

Query Match 79.1%; Score 17.4; DB 13; Length 355;
Best Local Similarity 94.7%; Pred. No. 3.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22
||| ||||||||||||||||
Db 136 CGAGCGGGGCGGGGCGAGC 154

RESULT 18
AW122220/c
LOCUS AW122220 375 bp mRNA linear EST 22-OCT-1999
DEFINITION UI-M-BH2.2-aov-d-07-0-UI.s1 NIH_BMAP_M_S3.2 Mus musculus cDNA clone UI-M-BH2.2-aov-d-07-0-UI 3', mRNA sequence.
ACCESSION AW122220
VERSION AW122220.1 GI:6097683
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 375)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined.
The following repetitive elements were found in this cDNA sequence:
192-247, >(CA)n#Simple_repeat
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers
source 1. :375
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH2.2-aov-d-07-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_M_S3.2"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S3.2 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S3.2, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S3.2) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S2 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the NIH_BMAP_M_S2 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S3.2 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S3.2
TAG_TISSUE=cerebellum
TAG_SEQ=GACTC"

BASE COUNT 87 a 136 c 77 g 75 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 9; Length 375;
Best Local Similarity 94.7%; Pred. No. 3.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22
||| ||||||||||||||||
Db 100 CGAGCGGGGCGGGGCGAGC 82

RESULT 19
AI846527/c
LOCUS AI846527 379 bp mRNA linear EST 15-JUL-1999
DEFINITION UI-M-AN1-aff-f-04-0-UI.s1 NIH_BMAP_MBG_N Mus musculus cDNA clone UI-M-AN1-aff-f-04-0-UI 3', mRNA sequence.
ACCESSION AI846527
VERSION AI846527.1 GI:5490433
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 379)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 192-247, >(CA)n#Simple_repeat
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers
source 1. .379
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AN1-aff-f-04-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_MBG_N"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MBG_N library is a normalized library constructed from mouse basal ganglia. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806 , 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
TAG_LIB=NIH_BMAP_MBG_N
TAG_TISSUE=hippocampus
TAG_SEQ=TTCGA"
BASE COUNT 89 a 136 c 77 g 77 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 9; Length 379;
Best Local Similarity 94.7%; Pred. No. 3.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22
||| ||||||||||||||||
Db 100 CGAGCGGGGCGGGGCGAGC 82

RESULT 20
AA615769
LOCUS AA615769 406 bp mRNA linear EST 07-OCT-1997
DEFINITION vo72f08.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone IMAGE:1064679 5' similar to WP:C16C10.7 CE01498 ZINC FINGER PROTEIN ;, mRNA sequence.
ACCESSION AA615769
VERSION AA615769.1 GI:2502997
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 406)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:587039
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 398.
FEATURES Location/Qualifiers
source 1. .406
/organism="Mus musculus"
/mol_type="mRNA"

```
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1064679"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead. The C2C12 cell line
                     (available from ATCC, catalog # CRL-1772) differentiates
                     rapidly, forming contractile myotubes and producing
                     characteristic muscle proteins. "
BASE COUNT       81 a     99 c    142 g     83 t      1 others
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 9;  Length 406;
  Best Local Similarity   94.7%;  Pred. No. 3.1e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        158 CGAGCGGGGCGGGGCGAGC 176


RESULT 21
CB765932
LOCUS       CB765932                418 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  AMGNNUC:MRPE4-00164-B12-A mrpe4 (10380) Rattus norvegicus cDNA
            clone mrpe4-00164-b12 5', mRNA sequence.
ACCESSION   CB765932
VERSION     CB765932.1  GI:29854323
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00164  row: b  column: 12.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="mrpe4-00164-b12"
                     /tissue_type="placenta embryo"
                     /clone_lib="mrpe4 (10380)"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI;
                     placenta embryo day 17"
BASE COUNT       89 a    108 c    144 g     70 t      7 others
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 14;  Length 418;
  Best Local Similarity   94.7%;  Pred. No. 3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        126 CGAGCGGGGCGGGGCGAGC 144


RESULT 22
BI220387
LOCUS       BI220387                423 bp    mRNA    linear   EST 11-JUL-2001
DEFINITION  602935686F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5098916 5',
            mRNA sequence.
ACCESSION   BI220387
VERSION     BI220387.1  GI:14673831
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11237  row: l  column: 21
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5098916"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT       88 a    104 c    155 g     76 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 12;  Length 423;
  Best Local Similarity   94.7%;  Pred. No. 3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        138 CGAGCGGGGCGGGGCGAGC 156


RESULT 23
BB859799
LOCUS       BB859799                437 bp    mRNA    linear   EST 26-NOV-2001
DEFINITION  BB859799 RIKEN full-length enriched, kidney CCL-142 RAG cDNA Mus
            musculus cDNA clone G430010A01 5', mRNA sequence.
ACCESSION   BB859799
VERSION     BB859799.1  GI:17101253
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
```

```
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .437
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="G430010A01"
                      /tissue_type="kidney"
                      /cell_line="CCL-142 RAG"
                      /clone_lib="RIKEN full-length enriched, kidney CCL-142 RAG
                      cDNA"
BASE COUNT       89 a    103 c    160 g     85 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 10;  Length 437;
  Best Local Similarity   94.7%;  Pred. No. 3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        168 CGAGCGGGGCGGGGCGAGC 186


RESULT 24
AI194235
LOCUS       AI194235                448 bp    mRNA    linear   EST 13-OCT-1998
DEFINITION  ue52e11.r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1494764 5' similar to TR:O35445 O35445 HYPOTHETICAL 19.8 KD
            PROTEIN. ;, mRNA sequence.
ACCESSION   AI194235
VERSION     AI194235.1  GI:3745442
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:932368
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .448
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:1494764"
                      /sex="female (lactating)"
                      /tissue_type="mammary gland"
                      /lab_host="DH10B"
                      /clone_lib="Soares_mammary_gland_NMLMG"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from mammary
                      gland tissue from a lactating female, and was then primed
                      with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                      ligated to Eco RI adaptors (Pharmacia), digested with Not
                      I  and cloned into the Not I and Eco RI sites of the
                      modified pT7T3 vector.  Library is normalized.  Library
                      was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       90 a    110 c    161 g     87 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 9;  Length 448;
  Best Local Similarity   94.7%;  Pred. No. 3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        143 CGAGCGGGGCGGGGCGAGC 161


RESULT 25
CB640552/c
LOCUS       CB640552                472 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEa15E16.f OSJNEa Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEa15E16 5', mRNA sequence.
ACCESSION   CB640552
VERSION     CB640552.1  GI:29635543
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: atc agc ggc cgc gat cc
            BACKWARD: aat taa ccc tca cta aag gg
```

```
           Plate: 15  row: E  column: 16
           Seq primer: atc agc ggc cgc gat cc.
FEATURES             Location/Qualifiers
     source          1. .472
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEa15E16"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEa"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 6 hrs after innoculation with Rice Blast (Che 86061
                     )"
BASE COUNT       79 a    146 c    166 g     81 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 14;  Length 472;
  Best Local Similarity   94.7%;  Pred. No. 3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 TCGATCGGGGCGGGGCGAG 21
              ||||||||||||| ||||||
Db         43 TCGATCGGGGCGAGGCGAG 25


RESULT 26
BE650594
LOCUS       BE650594                493 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  UI-M-BH2.2-aov-d-01-0-UI.r1 NIH_BMAP_M_S3.2 Mus musculus cDNA clone
            UI-M-BH2.2-aov-d-01-0-UI 5', mRNA sequence.
ACCESSION   BE650594
VERSION     BE650594.1  GI:9976418
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements The following repetitive elements were
            found in this cDNA sequence: 118-173, >(CA)n#Simple_repeat
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH2.2-aov-d-01-0-UI"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NIH_BMAP_M_S3.2"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S3.2 library is a subtracted library of a
                     series, ultimately derived from a mixture of individually
                     tagged normalized libraries from ten regions of the mouse
                     brain (cerebellum, brain stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal ganglia, pineal
                     gland, striatum, hipoccampus) after a series of
                     subtractions to reduce the representation of cDNAs from
                     which ESTs had already been generated. The following
                     serially subtracted libraries were generated in this
                     process: NIH_BMAP_M_S3.2, NIH_BMAP_M_S2, NIH_BMAP_M_S1.
                     The subtracted library (NIH_BMAP_M_S3.2) was constructed
                     as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S2
                     clones from which 3' ESTs had been derived was used as a
                     driver in a hybridization with the NIH_BMAP_M_S2 library
                     in the form of single-stranded circles. The remaining
                     single-stranded circles (subtracted library) was purified
                     by hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S3.2 library. This procedure has been
                     previously described (Bonaldo, Lennon and Soares, Genome
                     Research 6:791-806, 1996)"
BASE COUNT       99 a    112 c    158 g    124 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 10;  Length 493;
  Best Local Similarity   94.7%;  Pred. No. 3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| |||||||||||||||
Db        265 CGAGCGGGGCGGGGCGAGC 283


RESULT 27
BX529827
ID   BX529827   standard; RNA; EST; 523 BP.
XX
AC   BX529827;
XX
SV   BX529827.1
XX
DT   27-MAY-2003 (Rel. 75, Created)
DT   27-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   RZPD Mus musculus cDNA clone IMAGp952G154 = IMAGE:336469 5' EST.
XX
KW   EST; expressed sequence tag.
XX
OS   Mus musculus (house mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-523
RA   Heil O., Ebert L., Neubert P., Peters M., Radelof U., Schneider D.,
RA   Korn B.;
RT   ;
RL   Submitted (28-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer
RL   Feld 580, D-69120 Heidelberg, Germany
XX
CC   RZPD; IMAGp952G154.
CC   RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
CC   Mouse UnigeneSet - RZPD2 (RZPDLIB No.981)
CC   http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=981
CC   Contact: Ina Rolfs
CC   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
CC   Heubnerweg 6, D-14059 Berlin, Germany
CC   Tel: +49 30 32639 101
CC   Fax: +49 30 32639 111
CC   www.rzpd.de
```

```
CC   This clone is available royalty-free from RZPD;
CC   contact RZPD (clone@rzpd.de) for further information.
CC   Seq primer: T7, Primer sequence: TAATACGACTCACTATAGGG
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .523
FT                   /db_xref="taxon:10090"
FT                   /note="1st strand cDNA was primed with a Not I - oligo(dT)
FT                   primer [5'
FT                   TGTTACCAATCTGAAGTGGGAGCGGCCGCATGTTTTTTTTTTTTTTTTT 3'], on
FT                   total mouse RNA [provided by Minoru Ko, Wayne State Univ.];
FT                   double-stranded cDNA was ligated to EcoRI adaptors
FT                   5'-AATTCGGCACGAGG-3' and 5'-CCTCGTGCCG-3' (Pharmacia),
FT                   digested with NotI and cloned into the NotI and EcoRI sites
FT                   of the pT7T3D-PacI vector. Library went through one round
FT                   of normalization, and was constructed by Bento Soares and
FT                   M.Fatima Bonaldo."
FT                   /organism="Mus musculus"
FT                   /clone="IMAGp952G154"
FT                   /clone_lib="Soares 19.5dpc p3NMF19.5"
FT                   /dev_stage="19.5dpc total fetus"
FT                   /lab_host="DH10B"
XX
SQ   Sequence 523 BP; 110 A; 126 C; 186 G; 101 T; 0 other;

  Query Match             79.1%;  Score 17.4;  DB 4;  Length 523;
  Best Local Similarity   94.7%;  Pred. No. 2.9e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        237 CGAGCGGGGCGGGGCGAGC 255


RESULT 28
BF015640
LOCUS       BF015640                530 bp    mRNA    linear   EST 29-DEC-2000
DEFINITION  uy27b02.y1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:3660747 5'
            similar to TR:O35445 O35445 HYPOTHETICAL 19.8 KD PROTEIN. ;, mRNA
            sequence.
ACCESSION   BF015640
VERSION     BF015640.1  GI:10746972
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1421515
            Seq primer: -40RP from Gibco
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3660747"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Lu30"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      101 a    135 c    179 g    115 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 10;  Length 530;
  Best Local Similarity   94.7%;  Pred. No. 2.9e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        144 CGAGCGGGGCGGGGCGAGC 162


RESULT 29
BU919508
LOCUS       BU919508                600 bp    mRNA    linear   EST 30-OCT-2002
DEFINITION  6024-75 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
            mRNA sequence.
ACCESSION   BU919508
VERSION     BU919508.1  GI:24423360
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
            White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
  TITLE     Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
  JOURNAL   Nucleic Acids Res. 29 (24), 4983-4993 (2001)
  MEDLINE   21671825
   PUBMED   11812828
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT      114 a    161 c    197 g    128 t
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 13;  Length 600;
  Best Local Similarity   94.7%;  Pred. No. 2.9e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        134 CGAGCGGGGCGGGGCGAGC 152


RESULT 30
BU924221
LOCUS       BU924221                600 bp    mRNA    linear   EST 30-OCT-2002
DEFINITION  7082-67 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
            mRNA sequence.
ACCESSION   BU924221
VERSION     BU924221.1  GI:24428104
```

```
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
            White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
  TITLE     Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
  JOURNAL   Nucleic Acids Res. 29 (24), 4983-4993 (2001)
  MEDLINE   21671825
   PUBMED   11812828
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT      120 a    157 c    192 g    126 t      5 others
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 13;  Length 600;
  Best Local Similarity   94.7%;  Pred. No. 2.9e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        131 CGAGCGGGGCGGGGCGAGC 149


RESULT 31
CB578330
LOCUS       CB578330                624 bp    mRNA    linear   EST 03-APR-2003
DEFINITION  AMGNNUC:NRDG1-00162-E5-A nrdg1 (10855) Rattus norvegicus cDNA clone
            nrdg1-00162-e5 5', mRNA sequence.
ACCESSION   CB578330
VERSION     CB578330.1  GI:29522371
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00162  row: e  column: 5.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="nrdg1-00162-e5"
                     /tissue_type="Dorsal Root Ganglia"
                     /clone_lib="nrdg1 (10855)"
                     /note="Vector: pSPORT1; Site_1: Sal1; Site_2: Not1; rat
                     dorsal root ganglia"
BASE COUNT      113 a    174 c    201 g    135 t      1 others
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 14;  Length 624;
  Best Local Similarity   94.7%;  Pred. No. 2.9e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db        112 CGAGCGGGGCGGGGCGAGC 130


RESULT 32
BY721913
LOCUS       BY721913                635 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY721913 RIKEN full-length enriched, 12 days embryo, embryonic body
            between diaphragm region and neck Mus musculus cDNA clone
            9430090J12 5', mRNA sequence.
ACCESSION   BY721913
VERSION     BY721913.1  GI:27135030
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
```

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .635
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="9430090J12"
/tissue_type="embryonic body between diaphragm region and neck"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 12 days embryo, embryonic body between diaphragm region and neck"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI."

BASE COUNT 120 a 162 c 210 g 142 t 1 others
ORIGIN

Query Match 79.1%; Score 17.4; DB 14; Length 635;
Best Local Similarity 94.7%; Pred. No. 2.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22
||| ||||||||||||||||
Db 155 CGAGCGGGGCGGGGCGAGC 173

RESULT 33
BY721911
LOCUS BY721911 645 bp mRNA linear EST 17-DEC-2002
DEFINITION BY721911 RIKEN full-length enriched, 12 days embryo, embryonic body between diaphragm region and neck Mus musculus cDNA clone 9430089I02 5', mRNA sequence.
ACCESSION BY721911
VERSION BY721911.1 GI:27135028
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 645)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .645
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="9430089I02"
/tissue_type="embryonic body between diaphragm region and neck"
/dev_stage="12 days embryo"
/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, 12 days embryo, embryonic body between diaphragm region and neck"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI."
BASE COUNT 123 a 169 c 203 g 148 t 2 others
ORIGIN

Query Match 79.1%; Score 17.4; DB 14; Length 645;
Best Local Similarity 94.7%; Pred. No. 2.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22
||| ||||||||||||||||
Db 137 CGAGCGGGGCGGGGCGAGC 155

RESULT 34
BI111370
LOCUS BI111370 679 bp mRNA linear EST 26-JUN-2001
DEFINITION 602899252F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5028924 5', mRNA sequence.
ACCESSION BI111370
VERSION BI111370.1 GI:14562271
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 679)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11081 row: h column: 13
High quality sequence stop: 666.
FEATURES Location/Qualifiers
source 1. .679
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5028924"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam5"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"
BASE COUNT 126 a 190 c 216 g 146 t 1 others
ORIGIN

Query Match 79.1%; Score 17.4; DB 12; Length 679;
Best Local Similarity 94.7%; Pred. No. 2.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22
||| ||||||||||||||||
Db 129 CGAGCGGGGCGGGGCGAGC 147

RESULT 35
BZ658953
LOCUS BZ658953 704 bp DNA linear GSS 29-JAN-2003
DEFINITION OGAMF20TM ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0093C16, genomic survey sequence.
ACCESSION BZ658953
VERSION BZ658953.1 GI:28133843
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 704)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished
COMMENT Other_GSSs: OGAMF20TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .704
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMa0093C16"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
BASE COUNT 74 a 227 c 270 g 133 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 29; Length 704;
Best Local Similarity 94.7%; Pred. No. 2.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22
||| ||||||||||||||||
Db 660 CGACCGGGGCGGGGCGAGC 678

RESULT 36
BQ445767
LOCUS BQ445767 709 bp mRNA linear EST 29-MAY-2002
DEFINITION UI-M-ER0-bxm-g-18-0-UI.r2 NIH_BMAP_ER0 Mus musculus cDNA clone IMAGE:5710121 5', mRNA sequence.
ACCESSION BQ445767
VERSION BQ445767.1 GI:21248879
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```
REFERENCE   1  (bases 1 to 709)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .709
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5710121"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 15.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_ER0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site.  Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is GTGCGTGGAA. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      131 a    201 c    211 g    164 t      2 others
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 13;  Length 709;
  Best Local Similarity   94.7%;  Pred. No. 2.8e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db         90 CGAGCGGGGCGGGGCGAGC 108


RESULT 37
BQ444735
LOCUS       BQ444735                735 bp    mRNA    linear   EST 29-MAY-2002
DEFINITION  UI-M-ER0-bxm-h-21-0-UI.r1 NIH_BMAP_ER0 Mus musculus cDNA clone
            IMAGE:5710148 5', mRNA sequence.
ACCESSION   BQ444735
VERSION     BQ444735.1  GI:21247847
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5710148"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 15.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_ER0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site.  Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is GTGCGTGGAA. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      135 a    205 c    225 g    168 t      2 others
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 13;  Length 735;
  Best Local Similarity   94.7%;  Pred. No. 2.8e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| ||||||||||||||||
Db         90 CGAGCGGGGCGGGGCGAGC 108


RESULT 38
BQ445216
LOCUS       BQ445216                741 bp    mRNA    linear   EST 29-MAY-2002
DEFINITION  UI-M-ER0-bxp-d-17-0-UI.r1 NIH_BMAP_ER0 Mus musculus cDNA clone
            IMAGE:5711200 5', mRNA sequence.
ACCESSION   BQ445216
VERSION     BQ445216.1  GI:21248328
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
```

(BMAP)
Seq primer: pYX-5.
FEATURES Location/Qualifiers
source 1. .741
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5711200"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_ER0"
/note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNa was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail , is GTGCGTGGAA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Instututes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 133 a 202 c 224 g 179 t 3 others
ORIGIN

Query Match 79.1%; Score 17.4; DB 13; Length 741;
Best Local Similarity 94.7%; Pred. No. 2.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22
||| ||||||||||||||||
Db 129 CGAGCGGGGCGGGGCGAGC 147

RESULT 39
BI146925
LOCUS BI146925 764 bp mRNA linear EST 05-JUL-2001
DEFINITION 602911551F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5052660 5', mRNA sequence.
ACCESSION BI146925
VERSION BI146925.1 GI:14606926
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 764)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11143 row: e column: 13
High quality sequence stop: 741.
FEATURES Location/Qualifiers
source 1. .764
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5052660"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Li9"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 138 a 211 c 235 g 180 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 12; Length 764;
Best Local Similarity 94.7%; Pred. No. 2.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGATCGGGGCGGGGCGAGC 22
||| ||||||||||||||||
Db 115 CGAGCGGGGCGGGGCGAGC 133

RESULT 40
BY712338
LOCUS BY712338 773 bp mRNA linear EST 17-DEC-2002
DEFINITION BY712338 RIKEN full-length enriched, 10, 11 days embryo whole body Mus musculus cDNA clone 2810453M14 5', mRNA sequence.
ACCESSION BY712338
VERSION BY712338.1 GI:27123615
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 773)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda

,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .773
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2810453M14"
                     /tissue_type="whole body"
                     /dev_stage="10, 11 days embryo"
                     /clone_lib="RIKEN full-length enriched, 10, 11 days embryo
                     whole body"
BASE COUNT      144 a    204 c    238 g    186 t      1 others
ORIGIN

  Query Match             79.1%;  Score 17.4;  DB 14;  Length 773;
  Best Local Similarity   94.7%;  Pred. No. 2.8e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 CGATCGGGGCGGGGCGAGC 22
              ||| |||||||||||||||
Db        147 CGAGCGGGGCGGGGCGAGC 165


Search completed: February 18, 2004, 17:13:00
Job time : 1636 secs
```